Métodos de investigación del comportamiento

Octava edición

PAUL C. COZBY

California State University, Fullerton

Northcentral University

Traducción

Leticia Pineda Ayala

Lic. en Psicología

Universidad Anáhuac, Plantel Norte

Revisión técnica

Javier Urbina Soria

Nury Doménech Torrens

Raúl Ávila Santibáñez

Luis Emilio Cáceres Alvarado

Facultad de Psicología, UNAM

Se agradece la colaboración de Rocío Clavel Gómez y Sandra González López

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA
LISBOA • MADRID • NUEVA YORK • SAN JUAN • SANTIAGO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TORONTO

Gerente de división: Leonardo Newball González
Gerente editorial: Sergio Nájera Franco
Gerente de producto: Adolfo Bonilla Valdez
Editor de desarrollo: Carlos Edmundo Zúñiga Gutiérrez
Supervisor de producción: Timoteo Eliosa García

MÉTODOS DE INVESTIGACIÓN DEL COMPORTAMIENTO
Octava edición

Prohibida la reproducción total o parcial de esta obra,
por cualquier medio, sin autorización escrita del editor.

McGraw-Hill Interamericana

DERECHOS RESERVADOS © 2005 respecto a la primera edición en español por
McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V
*A Subsidiary of The **McGraw-Hill** Companies, Inc.*
Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A, Piso 17, Colonia Desarrollo Santa Fe, Delegación Alvaro Obregón 01376, México, D.F. Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN: 970-10-4825-3

Traducido de la primera edición en inglés de: METHODS IN BEHAVIORAL RESEARCH, EIGHT EDITION. Copyright © 2004 by The McGraw-Hill Companies, Inc. All rights reserved. Previous editions © 2001, 1997, 1993, 1989, 1985, 1981 by the Mayfield Publishing Company, © 1977 by Paul C. Cozby. ISBN: 007-2523425

1234567890

09876432105

Impreso en México

Printed in México

Esta obra se terminó de Imprimir en Octubre 2004 en Programas Educativos S.A. de C.V. Calz. Chabacano No 65-A Col. Asturias C.P 06850 Méx. D F Empresa certificada por el Instituto Mexicano de Normalización y Certificado A.C Bajo la Norma ISO-9002, 1994/NMX-CC-04 1995 con el num De registro RSC-048 y bajo la Norma ISO-14001 1996/SAA-1998 con el Núm. de Registro RSAA-003

Contenido

1 **COMPRENSIÓN CIENTÍFICA DEL COMPORTAMIENTO** 1

Usos de los métodos de investigación 2
La aproximación científica 3
Metas de la ciencia 8
Investigación básica y aplicada 11
Términos de estudio 14
Preguntas de repaso 14
Actividades 15

2 **POR DÓNDE EMPEZAR** 16

Hipótesis y predicciones 17
A quién estudiamos: una nota sobre los términos 18
Fuentes de ideas 18
Investigación bibliográfica 23
Anatomía de un artículo de investigación 34
Términos de estudio 36
Preguntas de repaso 36
Actividades 37

3 **ÉTICA DE LA INVESTIGACIÓN** 38

Experimento sobre obediencia de Milgram 39 El *Reporte Belmont* 40 Evaluación de riesgos y beneficios 40 Consentimiento informado 43 La importancia del informe 49 Alternativas al engaño 50 Justicia y la selección de los participantes 53 Compromisos del investigador 54 Reglamentos federales y el Consejo institucional de revisión 54

Código de ética de la APA 56
Investigación con participantes humanos 57
Ética e investigación con animales 60
Revisión de riesgos y beneficios 62
Falsificación 63
Términos de estudio 65
Preguntas de repaso 66
Actividades 66

4 **ESTUDIO DEL COMPORTAMIENTO** 69

Variables 70
Definiciones operacionales de variables 71
Relaciones entre variables 72
Métodos no experimentales *versus* experimentales 77
Variables independiente y dependiente 82
Causalidad 84
Elección de un método: ventajas de diversos métodos
Evaluación de la investigación: tres tipos de validez 89
Términos de estudio 91
Preguntas de repaso 91
Actividades 92

5 **CONCEPTOS SOBRE MEDICIÓN** 94

Confiabilidad de la medida 95
Validez de constructo de las mediciones 100
Reactividad de las mediciones 104
Variables y escalas de medición 105
Términos de estudio 108
Preguntas de repaso 108
Actividades 109

6 **OBSERVACIÓN DEL COMPORTAMIENTO** 11O

Métodos cuantitativo y cualitativo 111
Observación naturalista 112
Observación sistemática 116
Estudios de caso 119
Investigación documental 121
Descripción de la personalidad y diferencias individuales 124
Términos de estudio 125
Preguntas de repaso 126
Actividades 126

7 **PREGUNTAS HECHAS A LAS PERSONAS ACERCA DE SÍ MISMAS: INVESTIGACIÓN CON ENCUESTAS** 127

¿Por qué realizar encuestas? 128
Elaboración de preguntas 130
Respuestas a las preguntas 133
Finalización del cuestionario 137
Aplicación de encuestas 138
Diseños de encuestas para estudiar cambios a lo largo del tiempo 142
Muestreo a partir de una población 142
Técnicas de muestreo 145
Evaluación de muestras 148
Términos de estudio 151
Preguntas de repaso 151
Actividades 152

8 **DISEÑO EXPERIMENTAL** 153

Confusión y validez interna 154
Experimentos básicos 155
Asignación de los participantes a las condiciones experimentales 159
Diseño de grupos independientes 160
Diseño de medidas repetidas 160
Diseño de grupos apareados 160
Diseños de investigación del desarrollo 167
Términos de estudio 171
Preguntas de repaso 171
Actividades 172

9 **REALIZACIÓN DE EXPERIMENTOS** 174

Selección de los participantes de la investigación 175
Manipulación de la variable independiente 176
Medición de la variable dependiente 181
Controles adicionales 185
Consideraciones adicionales 190
Análisis e interpretación de resultados 193
Comunicar a otros la investigación 193
Términos de estudio 194
Preguntas de repaso 194
Actividades 195

10 **DISEÑOS EXPERIMENTALES COMPLEJOS** 197

Incremento del número de niveles de una variable independiente 198
Incremento del número de variables independientes: diseños factoriales 200
Términos de estudio 212
Preguntas de repaso 212
Actividades 213

11 **DISEÑOS CUASIEXPERIMENTALES Y EXPERIMENTALES DE CASO ÚNICO** 214

Evaluación de programas 215
Diseños cuasiexperimentales 218
Diseños experimentales de caso único 226
Términos de estudio 230
Preguntas de repaso 231
Actividades 231

12 **COMPRENSIÓN DE LOS RESULTADOS DE LA INVESTIGACIÓN: DESCRIPCIÓN Y CORRELACIÓN** 234

Escalas de medición: repaso 235
Análisis de los resultados de la investigación 236
Distribuciones de frecuencias 237
Estadística descriptiva 240
Gráficas de relaciones 241
Coeficientes de correlación: descripción de la fuerza de relaciones 243
Magnitud del efecto 247
Significación estadística 249
Ecuaciones de regresión 249
Correlación múltiple 250
Correlación parcial y el problema de la tercera variable 252
Modelos estructurales 253
Términos de estudio 254
Preguntas de repaso 254
Actividades 255

13 **COMPRENSIÓN DE LOS RESULTADOS DE INVESTIGACIÓN: INFERENCIA ESTADÍSTICA** 257

Muestras y poblaciones 258
Estadística inferencial 259
Hipótesis nula e hipótesis de investigación 259
Probabilidad y distribuciones muéstrales 260
Ejemplo: las pruebas *t* y *F* 263
Errores tipo I y tipo II 270
Elección de un nivel de significación 273
Interpretación de resultados no significativos 274
Elección de un tamaño de muestra: análisis de la potencia 276
La importancia de las replicaciones 277
Significación de un coeficiente r de correlación de Pearson 277
Análisis de datos por computadora 278
Selección de la prueba de significación adecuada 280
Términos de estudio 281
Preguntas de repaso 281
Actividades 282

14 **GENERALIZACIÓN DE LOS RESULTADOS** 284

Generalización a otras poblaciones 285
Consideraciones culturales 289
Generalización a otros experimentadores 290
Pretest y generalización 291
Generalización a partir de ambientes de laboratorio 292
La importancia de las replicaciones 293
Evaluación de la generalización mediante revisiones de la literatura y meta-análisis 296
Uso de la investigación para mejorar las condiciones de vida 298
Términos de estudio 299
Preguntas de repaso 299
Actividades 300

APÉNDICE A: REDACCIÓN DE REPORTES DE INVESTIGACIÓN 301

Introducción 301
Estilo de redacción 302
Organización del reporte 307
El uso de encabezados 317
Citas y fuentes de referencia 318
Abreviaturas 327
Algunas consideraciones gramaticales 329
Reporte de cifras y estadísticos 331
Conclusión 332
Ejemplo de un reporte 333

APÉNDICE B: PRUEBAS ESTADÍSTICAS 367

Estadística descriptiva 367
Pruebas de significación estadística 370
Correlación y magnitud del efecto 386

APÉNDICE C: TABLAS ESTADÍSTICAS 390
Tabla C.l Tabla de números aleatorios 392
Tabla C.2 Valores críticos de chi cuadrada 396
Tabla C.3 Valores críticos de *t* 397
Tabla C.4 Valores críticos de *F* 398
Tabla C.5 Valores críticos de *r* (coeficiente de correlación producto-momento de Pearson) 401

APÉNDICE D: CONSTRUCCIÓN DE UN CUADRADO LATINO 402
Glosario 404
Referencias 412
Créditos 424
Índice analítico 425

Prólogo a la edición en español

La psicología científica comparte características tanto de ciencias como la biología y la química, como de las ciencias sociales. Por lo tanto, los métodos de investigación que se emplean en psicología son comunes con los que se utilizan en otras áreas del conocimiento. Como profesión, la psicología tiene un vasto campo de aplicaciones y en todos los casos, desde las más tradicionales como en la psicología clínica y en la educativa hasta las más recientes como en la psicología ambiental y en la psicología de la política, se utilizan diversos métodos de investigación. Esta variedad de intereses en psicología ha resultado en una diversidad de textos que hacen énfasis en métodos de investigación apropiados para una u otra área de conocimiento en psicología. Sin embargo, en pocos textos se ofrece un panorama amplio de los diversos métodos de investigación ya sea de una manera didáctica y coloquial o con la profundidad apropiada en los temas relevantes.

En este libro, Paul C. Cozby presenta todo el panorama de métodos de investigación comúnmente empleados en psicología de una manera sencilla y práctica, con ejemplos de la vida cotidiana. La obra también incluye ejercicios fáciles de realizar de una manera autodidacta. Por estas razones impulsamos la traducción al español de este libro, que cubre de manera adecuada la gama de técnicas y procedimientos de método que se utilizan en la investigación básica y aplicada en psicología. Es un libro actualizado que incorpora recursos didácticos de vanguardia, como la indicación de los conceptos relevantes en cada capítulo, las preguntas de repaso, las actividades prácticas y un sitio de Internet en el que los alumnos pueden comprobar sus avances, resolver dudas y desarrollar ejercicios.

Esperamos que el libro resulte de utilidad como una introducción general a los métodos de investigación en psicología y que contribuya a la formación inicial de los estudiantes en el análisis científico del comportamiento humano.

Los revisores técnicos

Prefacio

La enseñanza y el aprendizaje de los métodos de investigación representan un reto, pero también es muy divertida. Esta nueva edición de *Métodos de investigación del comportamiento* mantiene las características de las ediciones anteriores que han sido apreciadas por los instructores y los estudiantes. Mi prioridad es una clara comunicación de los conceptos. He intentado presentar el material con claridad y utilizar ejemplos interesantes; también he tratado de incrementar el aprendizaje con la descripción de muchos conceptos importantes en diversos contextos a lo largo del libro; las investigaciones demuestran que la redundancia ayuda a la comprensión. También he hecho énfasis en la necesidad del estudio del comportamiento utilizando una variedad de aproximaciones de investigación. Cada capítulo se inicia con una introducción, al final de cada capítulo se incluyen términos de estudio, preguntas de repaso y actividades. Los términos importantes aparecen en negritas en el texto y se definen en el glosario.

Recursos

La página http://methods.fullerton.edu está dedicada al aprendizaje de los métodos de investigación. Este sitio ofrece una manera fácil para obtener más información acerca de los temas presentados en el texto, por medio de recursos disponibles en Internet.

Existe un Centro de Aprendizaje en Línea para estudiantes e instructores en www.mhhe.com/cozby8. Para los estudiantes, este recurso en línea brinda numerosos auxiliares de estudio para mejorar la experiencia del aprendizaje. Los estudiantes encontrarán preguntas de práctica, ejercicios en Internet y vínculos que complementan al libro de texto. Para los instructores, este recurso en línea incluye el *Manual del instructor,* un banco de imágenes y vínculos de Internet que sirven para preparar el material del curso. El *Manual del instructor* contiene muchas actividades y tareas para los estudiantes, así como preguntas de exámenes.

Organización

La organización generalmente sigue la secuencia de la planeación y realización de una investigación. Sin embargo, los capítulos son relativamente independientes para ofrecer a los instructores la máxima flexibilidad para determinar el or-

den de los capítulos. Por ejemplo, los capítulos sobre ética de investigación y métodos de investigación por encuestas se encuentran al inicio del libro, pero los instructores que deseen presentar este material posteriormente, pueden hacerlo con facilidad. El capítulo 1 ofrece un panorama general de la aproximación científica al conocimiento y establece las diferencias entre la investigación básica y la aplicada. El capítulo 2 se refiere a las fuentes de ideas de investigación y a la importancia de la investigación bibliográfica. El capítulo 3 se enfoca en la ética de la investigación; los aspectos éticos se estudian con profundidad aquí y se enfatizan a lo largo del libro. El capítulo 4 examina variables psicológicas y distingue entre las aproximaciones experimental y no experimental para el estudio de relaciones entre variables. El capítulo 5 cubre aspectos de la medición, incluyendo la confiabilidad y la validez. En el capítulo 6 se describen las aproximaciones de investigación no experimentales, incluyendo la observación naturalista, los estudios de caso y el análisis de contenido. En el capítulo 7 se estudia el muestreo, así como el diseño de cuestionarios y entrevistas. Los capítulos 8 y 9 presentan las bases del diseño y realización de experimentos. Los diseños factoriales se estudian en el capítulo 10. El capítulo 11 habla sobre las ventajas y desventajas de los diseños cuasiexperimentales y de caso único. Los capítulos 12 y 13 se enfocan en el uso de la estadística para comprender los resultados de las investigaciones. Finalmente, el capítulo 14 se refiere a aspectos de la generalización, el meta-análisis y la importancia de las replicaciones. También se incluyen apéndices sobre la redacción de los reportes de investigación, la elaboración de análisis estadísticos y la construcción de cuadrados latinos.

LO NUEVO DE ESTA EDICIÓN

Esta edición ha incorporado muchos cambios sugeridos por los usuarios de ediciones anteriores. El capítulo sobre ética de la investigación se ha reorganizado y ahora incluye los temas del *Reporte Belmont* y el nuevo Código de ética de la APA. En el apéndice A se presenta la quinta edición del *Manual de Publicaciones* de la APA. Además, se incluye un nuevo ejemplo de un artículo. La información estadística se ha ampliado, de modo que ahora se incluye material sobre la magnitud del efecto y los intervalos de confianza; también se ha añadido nuevo material gráfico en el capítulo sobre diseños experimentales complejos. Ahora todos los diseños cuasiexperimentales aparecen juntos en un solo capítulo.

RECONOCIMIENTOS

Muchas personas ayudaron a producir ésta y las ediciones anteriores del libro. Franklin Graham, quien guió este libro a lo largo de sus ediciones previas en Mayfield Publishing Company, se jubiló y se le extraña mucho. Por fortuna, Ken King de McGraw-Hill ha hecho un maravilloso trabajo como editor de esta edición. Agradezco a Adrienne Armstrong por la edición del manuscrito y a Hal

Lockwood de Penmarin Books por la gerencia de producción. Mitch Okada realizó un excelente trabajo al actualizar el *Manual del instructor, y* Diana Kyle hizo una increíble labor al crear el Centro de Aprendizaje en Línea. Estoy enormemente agradecido por la retroalimentación que he recibido de una gran cantidad de estudiantes y de instructores, y de forma particular agradezco a las siguientes personas que realizaron revisiones detalladas de esta edición:

Paul Atchley, University of Kansas
Lindsey L. Cohen, West Virginia University
Margaret M. Gittis, Youngstown State University
Merrill Hiscock, University of Houston
Diana Kyle, Fullerton College
Harold Sigall, University of Maryland, College Park
Ed de St. Aubin, Marquette University
Theodore E. Steiner, San Francisco State University

A modo de nota personal, me gustaría agradecer a las personas que forman parte de mi vida y que me han ayudado de diversas formas a producir el libro: Jeanne King, Josh Cozby, Brisco Cozby, Don Hecht, Gil Linne, David Perkins, Tom Klammer, Greg Robinson, Claire Palmerino, Dan Kee, Kathy Brown, Frank Bagrash, William Smith, Stan Woll, Dennis Berg, Bill Marelich, Kim Shattuck y Lisa Marr.

Estoy muy interesado en recibir comentarios y sugerencias de estudiantes e instructores. Por favor envíen un correo electrónico a cozby@fullerton.edu o ccozby@neu.edu

1

Comprensión científica del comportamiento

Usos de los métodos de investigación

La aproximación científica
Las limitaciones de la intuición y la autoridad
Autoridad
Escepticismo, ciencia y aproximación empírica
Integración de la intuición, el escepticismo y la autoridad

Metas de la ciencia
Descripción del comportamiento
Predicción del comportamiento
Determinación de las causas del comportamiento
Explicación del comportamiento

Investigación básica y aplicada
Investigación básica Investigación aplicada Comparación entre la investigación básica y la aplicada

Términos de estudio

Preguntas de repaso

Actividades

¿Cuáles son las causas de la agresión y la violencia? ¿De qué manera recordamos las cosas, qué nos hace olvidar y cómo podemos mejorar la memoria? ¿Cuáles son los efectos de los ambientes estresantes sobre la salud y la interacción social? ¿Cómo afectan las experiencias tempranas de la niñez en el desarrollo posterior? ¿Cuáles son las formas adecuadas para tratar la depresión? ¿Cómo podemos reducir los prejuicios y los conflictos entre los grupos? La curiosidad sobre preguntas como éstas tal vez sea la principal razón por la que muchos estudiantes deciden tomar cursos de ciencias de la conducta. La investigación científica nos ofrece un medio para plantearnos este tipo de preguntas y darles respuesta. En este libro examinaremos los métodos de investigación cien-tífica para las ciencias de la conducta. En el capítulo introductorio nos centra-remos en las formas en que el conocimiento de los métodos de investigación puede ser útil para comprender el mundo que nos rodea. Además, revisaremos las características de la aproximación científica al estudio del comportamiento y los temas generales de investigación que les interesan a los científicos de la conducta.

USOS DE LOS MÉTODOS DE INVESTIGACIÓN

De manera creciente, los ciudadanos que están informados en nuestra sociedad necesitan adquirir más conocimientos sobre los métodos de investigación. Los diarios, las revistas de interés general y otros medios de comunicación reportan resultados de investigaciones continuamente: "parece que los trastornos alimentarios son más comunes en lugares cálidos", "el ginkgo biloba mejora la memoria" o "el tabaquismo se relaciona con bajas calificaciones". Diversos artículos y libros versan sobre los efectos benéficos o dañinos de ciertas dietas o vitaminas en la vida sexual, la personalidad o la salud. Con frecuencia, se reportan resultados de encuestas que muestran conclusiones acerca de nuestras creencias con relación a una diversidad de temas. ¿Cómo evalúa usted estos reportes? ¿Acepta simplemente los hallazgos porque supuestamente son científicos? El conocimiento de los métodos de investigación le ayudará a leer estos reportes de forma crítica, a evaluar los métodos empleados y a decidir si las conclusiones son razonables.

Muchas ocupaciones requieren del uso de hallazgos obtenidos mediante investigaciones. Por ejemplo, los profesionales de la salud mental deben tomar decisiones acerca de métodos de tratamiento, sobre la asignación de pacientes a distintos tipos de instituciones, medicamentos y procedimientos de prueba. Es-tas decisiones deben tomarse con base en investigaciones; para tomar decisiones adecuadas, los profesionales de la salud mental deben ser capaces de revisar literatura científica en su campo y aplicarla en su vida profesional. De manera similar, las personas que trabajan en ambientes de negocios con frecuencia se basan en investigaciones para decidir sobre estrategias de mercado, formas para mejorar la productividad y el ánimo de los empleados, así como métodos para se

leccionar y entrenar a nuevos trabajadores. Los educadores deben mantenerse actualizados sobre la información de temas como la eficacia de distintas estrategias o programas de enseñanza, para tratar los problemas especiales de los estudiantes. El conocimiento de los métodos de investigación y la habilidad para evaluar reportes de investigación son útiles en muchos campos.

También es fundamental reconocer que la investigación científica se ha vuelto cada vez más importante en las decisiones de política pública. Los legisladores y líderes políticos en todos los ámbitos gubernamentales adoptan, con frecuencia, posiciones políticas y proponen leyes basadas en hallazgos de investigaciones. Las investigaciones también pueden influir en decisiones de tipo judicial: un excelente ejemplo de esto es el *Social Science Brief,* que fue preparado por psicólogos y aceptado como evidencia en el caso histórico de 1954 de *Brown vs. Board of Education,* en el que la Suprema Corte de Estados Unidos de América prohibió la segregación escolar en ese país. Uno de los estudios citados en el informe fue realizado por Clark y Clark (1947), en el cual se encontró que cuando se permite a los niños que elijan entre muñecas de piel clara y de piel oscura, tanto los niños blancos como los negros preferían jugar con las muñecas de piel clara (véase Stephan, 1983, para una mayor discusión sobre las implicaciones de este estudio). La legislación y la opinión pública acerca de la disponibilidad de materiales pornográficos han recibido información de las investigaciones sobre la conducta en este tema (véase, por ejemplo, Koop, 1987; Linz, Donnerstein y Penrod, 1987), y la investigación psicológica sobre los estereotipos sexuales ha influido enormemente en las decisiones de la Suprema Corte sobre la discriminación sexual de los empleadores (Fiske, Bersoff, Borgida, Deaux y Heilman, 1991). Además, los psicólogos que estudian formas para mejorar la identificación de testigos (por ejemplo, Wells *et al,* 1998; Wells, 2001) han ejercido una considerable influencia en los procedimientos de las investigaciones criminales, recomen-dados por las agencias dedicadas a la aplicación de la ley (Departamento de Justicia de Estados Unidos, 1999).

La investigación también es importante para la creación y evaluación de la eficacia de programas diseñados para alcanzar ciertas metas; por ejemplo, incrementar la permanencia de estudiantes en las escuelas, influir para que las personas realicen conductas que reduzcan su riesgo de contraer el VIH-SIDA, o permitir que los empleados de una compañía aprendan a reducir los efectos del estrés. Debemos ser capaces de determinar si estos programas tienen éxito en el cumplimiento de sus metas.

LA APROXIMACIÓN CIENTÍFICA

Abrimos este capítulo con varias preguntas sobre la conducta humana, y sugerimos que la investigación científica es un medio valioso para responderlas. ¿En qué difiere el método científico de otras formas de aprendizaje sobre el comportamiento? Las personas siempre han observado el mundo que las rodea y buscan

explicaciones para lo que ven y experimentan. Sin embargo, en lugar de utilizar el método científico, mucha gente confía en la intuición y la autoridad como formas de conocimiento.

Las limitaciones de la intuición y la autoridad

Muchos de nosotros sabemos o hemos escuchado sobre alguna pareja que, después de años de tratar de concebir, adopta un niño. Después de un corto periodo, la pareja descubre que la mujer está embarazada. Esta observación conduce a la creencia común de que la adopción incrementa la posibilidad de un embarazo en parejas que tienen dificultades para concebir un hijo. Esta conclusión parece ser razonable de manera intuitiva, y la gente suele dar una explicación para este efecto; por ejemplo, la adopción reduce una de las principales fuentes de estrés marital, y la disminución del estrés, a su vez, aumenta las posibilidades de concebir (véase Gilovich, 1991).

Este ejemplo ilustra el uso de la intuición y de las evidencias anecdóticas para sacar conclusiones generales sobre el mundo que nos rodea. Cuando usted se basa en la intuición, acepta sin ningún cuestionamiento que su propio juicio personal, o una historia aislada sobre la experiencia de otra persona, le dan información acerca del mundo. El método intuitivo toma varias formas. Con frecuencia, implica encontrar una explicación para nuestro propio comportamiento o el de otros. Por ejemplo, usted podría desarrollar una explicación del por qué de las persistentes desavenencias con un compañero de trabajo, tales como "que la otra persona desea mi cargo" o "tener que compartir el teléfono nos coloca en una situación conflictiva". En otras ocasiones, la intuición se utiliza para explicar sucesos intrigantes, como en el caso de concluir que la adopción aumenta las probabilidades de concebir en las parejas que tienen dificultades para tener hijos.

Uno de los problemas de la intuición es que diversos sesgos cognoscitivos y motivacionales afectan nuestras percepciones, por lo cual podemos sacar conclusiones erróneas sobre causas y efectos (por ejemplo, Fiske y Taylor, 1984; Nisbett y Ross, 1980; Nisbett y Wilson, 1977; Gilovich, 1991). Gilovich señala que de hecho no existe una relación entre la adopción y un embarazo subsecuente, según las investigaciones científicas. Entonces, ¿por qué mantenemos esta creencia? Lo más probable es que se deba a un sesgo cognoscitivo llamado *correlación espuria,* la cual ocurre cuando nos enfocamos en dos sucesos que aparecen y suceden juntos. Cuando una adopción es seguida casi de inmediato por un embarazo, la situación atrae nuestra atención y tendemos a concluir que debe existir una conexión causal. Este tipo de correlaciones espurias también suelen ocurrir cuando tenemos una alta motivación para creer en la relación causal. Aun cuando esto es natural en nosotros, no es científico. El método científico requiere mucha más evidencia antes de sacar conclusiones.

Autoridad

El filósofo Aristóteles se interesó por los factores asociados con la persuasión o cambio de actitudes. En su *Retórica,* Aristóteles describe la relación entre la persuasión y la credibilidad: "la persuasión se logra por medio del carácter personal del orador, cuando el discurso se da de modo que pensemos que es creíble. Creemos más y con mayor facilidad en los hombres buenos que en otros". Por lo tanto, Aristóteles diría que tenemos mayores posibilidades de ser persuadidos por un orador que nos parece más prestigioso, confiable y respetable, que por uno que carece de estas cualidades.

Muchos de nosotros podríamos aceptar los argumentos de Aristóteles, sencillamente porque está considerado como una "autoridad" prestigiosa y porque sus escritos continúan siendo importantes. De forma similar, muchas personas están dispuestas a aceptar de antemano cualquier cosa que aprendan de los noticiarios, libros, funcionarios del gobierno o figuras religiosas. Consideran que las afirmaciones de estas autoridades deben ser verdaderas. El problema, por supuesto, es que probablemente las afirmaciones no sean verdaderas. El mé-:odo científico rechaza la idea de aceptar las aseveraciones de cualquier auto-ridad *por fe;* nuevamente, se requiere de mayor evidencia antes de poder sacar conclusiones científicas.

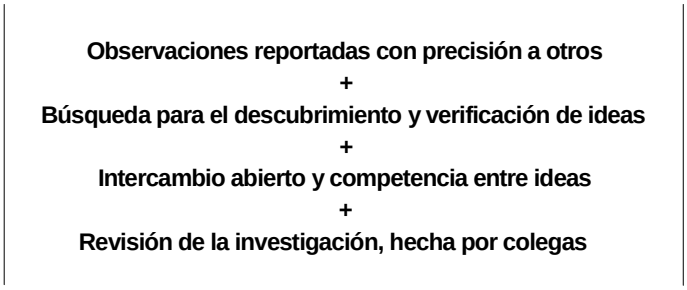
Escepticismo, ciencia y aproximación empírica

La aproximación científica del conocimiento reconoce que la intuición y la autoridad son fuentes de ideas sobre el comportamiento. Sin embargo, los científicos no aceptan, sin un cuestionamiento de por medio, la intuición de cualquiera, incluso, la suya. Los científicos reconocen que sus ideas tienen las mismas posibilidades de ser incorrectas que las de cualquier otro. Además, los científicos no aceptan por fe los pronunciamientos de cualquiera, independientemente del prestigio o autoridad de esta persona. Por lo tanto, los científicos son muy escépticos respecto con lo que ven y escuchan. El escepticismo científico implica que las ideas deben evaluarse con base en la lógica y en resultados cuidadosos de investigaciones científicas.

Si los científicos rechazan la intuición y la aceptación ciega de la autoridad como formas de conocimiento sobre el mundo, ¿cómo es que obtienen conocimientos? La característica fundamental del método científico es el **empirismo;** es decir, el conocimiento está basado en las observaciones. La recolección de da-tos forma la base de las conclusiones sobre la naturaleza del mundo. El método científico comprende una serie de reglas para la recolección y evaluación de da-tos; se explorarán estas reglas a lo largo del libro.

El poder del método científico puede observarse a nuestro alrededor. Ya sea en biología, química, medicina, física, antropología o psicología, se observan

FIGURA 1.1
Elementos de la teoría desarrollada de la ciencia, de acuerdo con Goodstein.



avances sorprendentes durante los pasados 25, 50 o 100 años. Tenemos una mayor comprensión del mundo que nos rodea, y ya se ha llevado a cabo la aplicación de tal entendimiento. Goodstein (2000) describe una "teoría desarrollada de la ciencia" que define las características de la indagación científica, éstas se resumen en la figura 1.1.

La primera explica que los científicos realizan observaciones que reportan con precisión a otros científicos y al público, otros individuos pueden replicar los métodos utilizados y obtener los mismos resultados. Además, la falsificación de datos implica, de manera inherente, una falta de ética y merece fuertes sanciones. En segundo lugar, los científicos buscan con entusiasmo observaciones para verificar sus ideas sobre el mundo; desarrollan teorías, argumentan qué datos existentes apoyan sus teorías y realizan investigaciones que puedan aumentar nuestra confianza en que las teorías son correctas. En tercer lugar, la ciencia florece cuando existe un sistema abierto para el intercambio de ideas. Se puede realizar investigación para probar cualquier idea que se esté planteando; quienes apoyan la idea y quienes están en desacuerdo con ella pueden aportar sus hallazgos de investigación, los cuales son susceptibles de ser evaluados por otras personas. Algunas ideas, incluso muy buenas, pueden resultar falsas; la investigación no logra proporcionar apoyo para ellas. Cuando esto ocurre, también avanza la ciencia, ya que puede provocar el desarrollo de nuevas y mejores ideas. Finalmente, la revisión que los colegas hacen de las investigaciones es importante para asegurar que sólo se publiquen las mejores de ellas. Antes de que un estudio se publique en una revista científica, debe ser revisado por colegas. Esto asegura que las investigaciones con grandes fallas no se conviertan en parte de la literatura científica. En esencia, la ciencia existe en un mercado libre de ideas, en el que las mejores son apoyadas por la investigación, y los científicos pueden construir con base en los estudios de otros profesionales para lograr mayores avances.

Integración de la intuición, el escepticismo y la autoridad

La ventaja del método científico sobre otras formas de conocimiento del mundo radica en que proporciona un conjunto objetivo de reglas para reunir, evaluar y reportar información. Se trata de un sistema abierto que permite que las ideas sean rechazadas o apoyadas por los demás. No obstante, esto no significa que la intuición y la autoridad no sean importantes. Como se señaló antes, con frecuencia los científicos confían en la intuición y en las aseveraciones de autoridades para obtener ideas de investigación. Además, no hay nada de malo en aceptar las afirmaciones de una autoridad, siempre y cuando no las aceptemos como evidencia científica. Con frecuencia no es posible obtener evidencia cien-tífica, como sucede, por ejemplo, cuando las religiones nos piden aceptar ciertas creencias por fe. Algunas creencias no pueden ser probadas y, por lo tanto, se encuentran fuera del campo de la ciencia. Sin embargo, en la ciencia las ideas de-ben evaluarse sobre la base de evidencias disponibles que puedan ser utilizadas para apoyarlas o rechazarlas.

Tampoco hay nada de malo en tener opiniones o creencias, siempre y cuan-do se presenten simplemente como tales. Sin embargo, debemos preguntar si la opinión puede probarse científicamente o si existe evidencia científica que se relacionen con ella. Por ejemplo, las opiniones respecto a un incremento en la agresión por la exposición a la violencia televisiva continúan siendo opiniones hasta que se reúna evidencia científica sobre el tema.

Conforme aprenda más sobre los métodos científicos, usted se volverá cada vez más escéptico de los resultados de investigaciones reportadas en los medios de comunicación, y también de las afirmaciones de los científicos. Usted debe estar consciente de que los científicos, con frecuencia, se convierten en autoridades cuando expresan sus ideas. Cuando alguien afirma ser un científico, ¿debemos estar más dispuestos a aceptar lo que dice? Primero pregunte sobre las referencias de esta persona. Tendemos a prestar mayor atención a alguien que tiene una reputación establecida en el campo; también nos vemos influenciados por el prestigio de la institución que esta persona representa. Asimismo, vale la pena examinar quién financia al investigador; por ejemplo, uno podría tener cierta suspicacia si la investigación costeada por una compañía farmacológica apoya la eficacia de un medicamento fabricado por dicha empresa. De manera similar, si una organización con actividades sociopolíticas financia investigaciones que apoyan esos fines, tendemos a mostrarnos escépticos de los hallazgos y a examinar de manera detallada los métodos utilizados en el estudio. Además, existen muchos "seudocientíficos" que emplean términos científicos para dar mayor peso a sus afirmaciones; éstos incluyen desde astrólogos hasta mercadólogos,

quienes desean que usted compre productos para mejorar su memoria o aumentar su impulso sexual. Finalmente, todos somos susceptibles a los reportes de hallazgos científicos falsos que circulan a través de Internet. Muchos de los reportes suelen estar asociados con un científico u organización científica de gran reputación, y después toman vida propia. Un reporte reciente, que recibió una amplia cobertura y que supuestamente provenía de la Organización Mundial de la Salud, afirmaba que el gen del cabello rubio estaba siendo eliminado del mapa genético; ¡el cabello rubio sería un rasgo en extinción! Una regla general es ser escéptico cuando se hacen afirmaciones científicas que están apoyadas únicamente en evidencias vagas o improbables.

METAS DE LA CIENCIA

La investigación científica tiene cuatro metas generales: 1) describir el comportamiento; 2) predecir el comportamiento; 3) determinar las causas del comportamiento, y 4) comprender o explicar el comportamiento.

Descripción del comportamiento

Ya que la primera meta de la ciencia es describir sucesos, el científico inicia con una observación cuidadosa. Cunningham y sus colegas examinaron juicios acerca del atractivo físico a lo largo del tiempo (Cunningham, Druen y Barbee, 1997). En 1976, estudiantes universitarios varones calificaron el atractivo de un gran número de mujeres presentadas en fotografías. En 1993, otro grupo de estudiantes evaluó las mismas fotografías. Los juicios sobre el atractivo de las mujeres fueron virtualmente idénticos; los estándares de la atracción aparentemente cambiaron muy poco durante este tiempo. En otro estudio, Cunningham comparó las características faciales de mujeres que eran estrellas de cine durante los años treinta y cuarenta, con las de artistas de los años noventa. Las comparaciones incluyeron, entre otras, el ancho y largo de los ojos, la longitud de la nariz, la prominencia de las mejillas y lo ancho de la sonrisa. Estas características faciales se asemejaban mucho en los dos periodos, indicando nuevamente que los estándares del atractivo permanecen constantes al paso del tiempo.

Con frecuencia, a los investigadores les interesa describir las formas en que los acontecimientos se relacionan sistemáticamente entre sí. ¿Procesan los jueces con mayor benevolencia a los acusados que son atractivos que a aquellos que no lo son? ¿Tiende más la gente a ser persuadida por un orador con alta credibilidad? ¿De qué manera cambian las habilidades cognoscitivas cuando las personas crecen? ¿Obtienen menores calificaciones en los exámenes los alumnos que estudian con el televisor encendido, que los estudiantes que lo hacen en un ambiente silencioso?

Predicción del comportamiento

Otra meta de la ciencia es predecir el comportamiento. Una vez que se ha observado con cierta regularidad que dos hechos están relacionados sistemáticamente entre sí (por ejemplo, una mayor credibilidad está asociada con mayores cambios en las actitudes), entonces es posible hacer predicciones. Una implicación de esto es la posibilidad de anticipar sucesos. Si sabemos que un candidato en una elección tiene mayor credibilidad que otro, podremos ser capaces de anticipar los resultados de la elección. Además, la posibilidad de pronosticar suele ayudar a tomar mejores decisiones. Por ejemplo, muchos estudiantes universitarios miden sus intereses vocacionales con pruebas como el Inventario de intereses Strong-Campbell, en el centro de orientación vocacional universitario, ya que el conocimiento de sus puntuaciones puede ayudarlos a tomar mejores decisiones sobre sus metas profesionales y la elección de una carrera.

Determinación de las causas del comportamiento

La tercera meta de la ciencia es determinar las causas del comportamiento. Aun cuando podamos predecir con precisión la ocurrencia de una conducta, es probable que no hayamos identificado sus causas. Por ejemplo, las puntuaciones en las pruebas de aptitudes no causan las calificaciones universitarias. La prueba de aptitudes es un indicador de otros factores que son las causas reales; se pueden realizar investigaciones para estudiar estos elementos. De manera similar, los estudios han mostrado que la conducta agresiva de un niño puede predecirse si se sabe cuánta violencia observa en la televisión. Desgraciadamente, a menos que sepamos que la exposición a la violencia televisiva es una *causa* de la conducta, no podemos afirmar que el comportamiento agresivo puede reducirse con disminuir las escenas violentas de la televisión. Por consiguiente, para saber cómo *cambiar* la conducta, es necesario conocer las *causas* de sus orígenes. La identificación de una de las causas de cierta conducta requiere de tres tipos de evidencias. No es suficiente con saber que dos sucesos ocurren juntos, como cuando sabemos que observar violencia en la televisión permite pronosticar los niveles reales de agresión. Para determinar una causa, deben ocurrir tres cosas:

1. Existe un orden temporal de los acontecimientos, en el que la causa precede al efecto. Esto se denomina **precedencia temporal.** Por lo tanto, necesitamos saber que primero se observó violencia televisiva y que luego se dio la agresión.
2. Cuando la causa está presente, ocurre el efecto; cuando la causa no está presente, el efecto no se presenta. A esto se le llama **covariación de la causa y el efecto.** Necesitamos saber que los niños que ven violencia en la

televisión se comportan de forma agresiva, y que los niños que no ven violencia en la televisión no se conducen de esa manera.

3. Ninguna otra variable causal podría ser responsable del efecto observado. Esto se *conoce* como eliminación de **explicaciones alternativas.** No deben existir otras explicaciones alternativas posibles para la relación. Este tercer punto es muy importante; suponga que a los niños que ven una cantidad considerable de violencia en la televisión se les deja más tiempo solos que a los niños que no ven este tipo de violencia. En este caso, el incremento en la agresión podría tener una explicación alternativa; la falta de supervisión de los padres. La causalidad se discutirá nuevamente en capítulos posteriores.

Explicación del comportamiento

La meta final de la ciencia es explicar los hechos que ya han sido descritos. El científico busca comprender *por qué* sucede la conducta. Considere la relación entre la violencia televisiva y la agresión, aun cuando sepamos que la violencia televisiva es una causa de la agresividad, es necesario explicar la relación. ¿Se debe a la imitación o "modelamiento" de la violencia observada en la televisión? ¿Es el resultado de la insensibilización psicológica ante la violencia y sus efectos? ¿O la violencia televisiva conduce a la creencia de que la agresión es una respuesta normal a la frustración y al conflicto? Se requiere de más investigaciones que brinden mayor información sobre posibles explicaciones de lo que se ha observado. Por lo general, estudios adicionales como éstos se realizan por medio de la prueba de teorías que se elaboran para explicar conductas específicas.

La descripción, predicción, determinación de las causas y la explicación están interrelacionadas estrechamente. La determinación de los motivos y las explicaciones del comportamiento están particularmente vinculadas, ya que es difícil conocer, incluso, la verdadera fuente o todos los móviles de cualquier conducta. Una explicación que parece ser satisfactoria se puede volver inadecuada cuando se identifican otras causas en investigaciones subsecuentes. Por ejemplo, cuando estudios pioneros mostraron que la credibilidad de un orador está asociada con el cambio de actitudes, los investigadores explicaron el hallazgo al afirmar que las personas tienden más a aceptar lo dicho por un individuo con una alta credibilidad, que por uno que tiene poco crédito. Sin embargo, esta explicación ha dado paso a una teoría más compleja sobre el cambio de actitudes que toma en cuenta muchos otros factores que están relacionados con la persuasión (Petty y Cacioppo, 1986). En resumen, existe cierta ambigüedad en la empresa de la búsqueda científica. Los nuevos descubrimientos de la investigación casi siempre hacen surgir nuevas preguntas que deben ser retomadas en otros estudios; con frecuencia, las explicaciones del comportamiento deben descartarse o revisarse conforme se reúnen nuevas evidencias. Esta ambigüedad forma parte de lo apasionante y divertido de la ciencia.

INVESTIGACIÓN BÁSICA Y APLICADA

Investigación básica

La investigación básica trata de responder preguntas fundamentales sobre la naturaleza del comportamiento. De manera frecuente, los estudios se diseñan para tratar aspectos teóricos sobre fenómenos como la cognición, la emoción, la motivación, el aprendizaje, la psicobiología, el desarrollo de la personalidad y la conducta social. A continuación se presentan las citas de algunos artículos científicos que abordan preguntas de investigación básica:

Marian, V. y Neisser, U. (2000). Language dependent recall of autobiographical memories. *Journal of Experimental Psycbology: General, 129,* 361-368.

Los individuos bilingües que hablan ruso e inglés recordaron más información sobre el periodo de sus vidas en el que hablaban ruso, cuando fueron entrevistados en ese idioma. Las mismas personas recordaron más de las épocas de su vida en que hablaban inglés, cuando se les entrevistó en inglés.

Watkins, M. J., LeCompte, D. C. *y* Kyungmi, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psycbology: Learning Memory and Cognition, 26,* 239-245.

Cuando se les pidió a los participantes que recordaran palabras comunes y raras, presentadas en listas separadas, recordaron un mayor número de las comunes. No obstante, si las mismas palabras se presentaban en listas mezcladas, los participantes recordaban más palabras raras; cuando se presentan juntas, aparentemente la estrategia de estudio es enfocarse en las palabras inusuales.

Raine, A., Reynolds, C, Venables, P. H. y Mednick, S. A. (2002). Stimulation seeking and intelligence: A prospective longitudinal study. *Journal of Personality and Social Psycbology, 82,* 663-674.

Los niños de tres años de edad que buscaron mayor estimulación presentaron puntuaciones más altas en pruebas de inteligencia a los 11 años, que los infantes de tres años que buscaron menor estimulación. Los niños dispuestos a buscar mayor estimulación quizás crean ambientes con mayor incitación cognoscitiva, lo que a su vez incrementa sus habilidades cognoscitivas.

Investigación aplicada

Los artículos de investigación citados anteriormente se refieren a procesos básicos del comportamiento y la cognición, y no tienen implicaciones prácticas

inmediatas. En contraste, la investigación aplicada se realiza para tratar temas en los que existen problemas prácticos y soluciones potenciales. Como ejemplo, a continuación se incluyen algunos títulos de artículos científicos:

Wells, M. y Perrine, R. (2001). Critters in the cube farm: Perceived psychological and organizational effects of pets in the workplace. *Journal of Occupational Health Psycbology,* 6, 81-87.

Se pidió a personas que dirigen empresas que permiten a sus trabajadores llevar a sus mascotas, que contaran sus experiencias. En general, los individuos consideraron que la presencia de las mascotas reduce el estrés del trabajo y que beneficia a la organización; este efecto fue más fuerte entre los empleados que llevaban mascotas, que entre quienes no las llevaban o no tenían.

Bushman, B. J. y Bonacci, A. M. (2002). Violence and sex impair memory for television ads. *Journal of Applied Psychology, 87,* 557-564.

Los participantes vieron un programa de televisión que contenía violencia, otro con escenas sexuales u otro neutral sin violencia ni sexo. Los programas incluyeron los mismos anuncios. Se estableció comunicación telefónica con los participantes al día siguiente para medir su recuerdo de los anuncios. Los asistentes participantes en la condición con contenido neutral recordaron más anuncios que aquellos que vieron los programas con contenido sexual o violento.

Straughan, R. D. y Lynn, M. (2002). The effects of salesperson compensation on perceptions of salesperson honesty. *Journal of Applied Social Psycbology, 32,*719-731.

Los participantes calificaron la honestidad de hombres y mujeres vendedores que trabajaban por salario o por comisiones. Los vendedores que laboraban por comisiones fueron juzgados menos honestos que los que lo hacían por salario. Sin embargo, la compensación no afectó las calificaciones de honestidad de las vendedoras.

Una de las áreas vitales de la investigación aplicada es la llamada **evaluación de programas.** La investigación de evaluación de programas califica las reformas e innovaciones sociales que se dan en el gobierno, la educación, el sistema de justicia criminal, la industria, el cuidado de la salud y las instituciones de salud mental. En un artículo de mucha influencia sobre "las reformas como experimentos", Campbell (1969) observó que los programas sociales en realidad son experimentos diseñados para lograr ciertos resultados. Afirmó, de manera persuasiva, que los científicos sociales deben evaluar cada programa para determinar si está logrando los efectos esperados. Si no es así, se deben probar programas alternativos. Este es un punto importante que con frecuencia las personas en todas las organizaciones no recuerdan cuando se aplican nuevas ideas; el

método científico dicta que deben evaluarse nuevos programas. Aquí se muestran dos ejemplos de artículos científicos:

Grossman, J. B. y Tierney, J. P. (1998). Does mentoring work? An impact study of the Big Brothers Big Sisters program. *Evaluation Review, 22,* 403-426.

Se *realizó* un experimento para evaluar el efecto de la participación en el programa *Big Brothers Big Sisters.* Los jóvenes de entre diez y 16 años de edad, participantes en el programa, manifestaron menores posibilidades de faltar a la escuela, de iniciar el consumo de drogas o alcohol o de verse involucrados en peleas, que los jóvenes del grupo control.

Celio, A. A., Winzelberg, A. J., Dev, P. y Taylor, C. B. (2002). Improving compliance in on-line, structured self-help programs: Evaluation of an eating disorder prevention program. *Journal of Psychiatric Practice, 8,* 14-20.

Mujeres universitarias utilizaron uno de cuatro distintos programas de autoayuda en computadora para reducir trastornos alimentarios. Un programa que utiliza un método estructurado, con recordatorios por correo electrónico, se relacionó con una mayor lectura de las pantallas de programas y una progresiva mejoría de la conducta.

Gran parte de la investigación aplicada se realiza en ambientes como gran-des empresas, compañías de investigación de mercado, agencias del gobierno y organizaciones de encuestas públicas; éstas no se publican, sino que se utilizan dentro de la empresa o por los clientes de ésta. No obstante, ya sea que los resultados se publiquen o no, se utilizan para ayudar a las personas a tomar mejores decisiones con respecto de problemas que requieren de una acción inmediata.

Comparación entre la investigación básica y la aplicada

Tanto la investigación básica como la aplicada son importantes, y ninguna de ellas puede considerarse superior a la otra. De hecho, el progreso de la ciencia depende de una sinergia entre la investigación básica y la aplicada. Gran parte de la investigación aplicada está guiada por las teorías y los hallazgos de investigaciones básicas. Por ejemplo, la investigación aplicada sobre el testimonio de expertos en juicios está guiada por la investigación básica en percepción y cognición. A su vez, los hallazgos obtenidos en los ambientes aplicados con frecuencia requieren de la modificación de teorías existentes y obligan a mayor investigación básica. Así, el estudio del testimonio de testigos reales produce un conocimiento más rico y preciso de los procesos perceptivos y cognoscitivos básicos.

En años recientes, muchos individuos en nuestra sociedad, incluyendo legisladores que controlan el presupuesto gubernamental de las agencias que financian las investigaciones, han exigido que los estudios sean pertinentes de manera directa para aspectos sociales específicos. El problema con esta actitud hacia la investigación es que nunca podemos predecir las aplicaciones últimas de la

investigación básica. Por ejemplo, en los años treinta el psicólogo B. F. Skinner realizó investigación básica sobre el condicionamiento operante, el cual describió cuidadosamente los efectos del reforzamiento en conductas como una rata que presiona una palanca. Años después, estas investigaciones propiciaron mayores aplicaciones prácticas en terapia, educación y psicología industrial. Los estudios que no tienen un valor *práctico* aparente, después se pueden volver muy útiles. El hecho de que nadie pueda prever la trascendencia eventual de la investigación básica, conduce a la conclusión de que el apoyo a este tipo de estudios es necesario tanto para el avance de la ciencia como para el beneficio de la sociedad.

La investigación del comportamiento es importante en muchos campos y tiene aplicaciones en las políticas públicas. Este capítulo lo ha introducido a usted a las principales metas y a los tipos generales de investigación. Todos los investigadores emplean métodos científicos, ya sea que estén interesados en asuntos básicos, aplicados o de evaluación de programas. Los temas y conceptos de *este* capítulo se ampliarán en el resto del libro; serán la base con la que usted evaluará la investigación de otros individuos y con la que planeará sus propios proyectos de investigación.

En este capítulo se enfatizó que los científicos son escépticos sobre lo que es verdadero en el mundo, insisten en que las proposiciones sean probadas de manera empírica. En los siguientes dos capítulos, nos enfocaremos en otras *dos* características de los científicos. Primero, los científicos sienten una intensa curiosidad sobre el mundo y encuentran inspiración para nuevas ideas en muchos lugares. En segundo lugar, tienen fuertes principios éticos, están comprometidos a tratar a quienes participan en las investigaciones con respeto y dignidad.

Términos de estudio

Autoridad
Covariación de causa y efecto
Empirismo
Escepticismo
Evaluación de programas
Explicaciones alternativas
Intuición
Investigación aplicada
Investigación básica
Metas de la ciencia
Precedencia temporal

Preguntas de repaso

1. ¿Por qué es importante que todos los individuos en nuestra sociedad conozcan los métodos de investigación?
2. ¿Por qué el escepticismo científico sirve para incrementar nuestro conocimiento del comportamiento? ¿En qué difiere el método científico de otras formas de búsqueda de conocimiento sobre el comportamiento?
3. Aporte definiciones y ejemplos de las metas de la investigación científica: descripción, predicción, determinación de las causas y explicación.

4. ¿Cuáles son las diferencias entre la investigación básica y la investigación aplicada?
5. Describa las características de la forma en que funciona la ciencia, según Goodstein (2000).

Actividades

i. Lea varios artículos de un periódico e identifique las fuentes utilizadas para apoyar las afirmaciones y conclusiones. ¿Utilizó el escritor la intuición, apeló a la autoridad, a la evidencia científica o a una combinación de éstas? Dé ejemplos específicos.

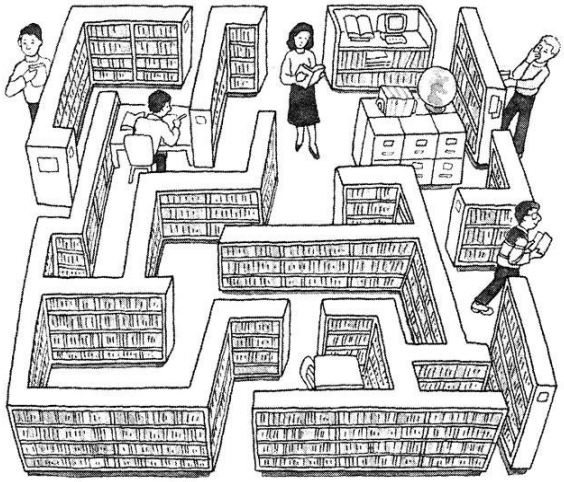
2. Imagine un debate sobre la siguiente afirmación: los científicos del comportamiento únicamente deben realizar investigaciones que tengan aplicaciones prácticas inmediatas. Elabore argumentos que apoyen (en favor) y desmientan (en contra) la afirmación.

3. Imagine un debate sobre la siguiente aseveración: el conocimiento de los métodos de investigación es innecesario para los estudiantes que tienen la intención de estudiar una carrera en psicología clínica. Elabore argumentos que apoyen (en favor) y desmientan (en contra) la afirmación.

4. El encabezado de un periódico dice que: "los trastornos alimentarios quizás sean más comunes en lugares cálidos". Usted lee el artículo y descubre que un investigador encontró que la incidencia de los trastornos alimentarios entre mujeres estudiantes de una universidad de Florida, era más alta que en una universidad de Pennsylvania. Suponga que en este estudio se describe con precisión la diferencia entre los estudiantes en las dos universidades. Discuta los hallazgos en términos de la identificación de causa y efecto y la explicación. Regrese a esta pregunta después de que lea los próximos capítulos. Para obtener mayor información, vea Sloan, D. M. (2002). Does warm weather climate affect eating disorder pathology? *International Journal of Eating Disorders, 32,* 240-244.

5. Identifique formas en las que usted podría haberse permitido aceptar creencias o involucrarse en prácticas que habría rechazado si hubiera tomado una postura de escepticismo científico. Por ejemplo, deberíamos recordar que una aseveración hecha en un correo electrónico puede ser un engaño o rumor. Ofrezca detalles específicos de la(s) experiencia(s). ¿Qué haría para investigar si la aseveración es válida?

2



Por dónde empezar

Hipótesis y predicciones

A quién estudiamos: una nota sobre los términos

Fuentes de ideas
El sentido común
La observación del mundo que nos rodea
Teorías
Investigaciones previas
Problemas prácticos

Investigación bibliográfica
La naturaleza de las revistas científicas
Psychological Abstracts (resúmenes psicológicos)
Cómo hacer una búsqueda en *PsycINFO*
Science Citation Index y *Social Science Citation Index*
Revisiones de la literatura
Otros recursos electrónicos de búsqueda
Búsquedas en Internet

Anatomía de un artículo de investigación
Resumen
Introducción
Método
Resultados
Discusión

Términos de estudio

Preguntas de repaso

Actividades

La motivación para realizar investigación científica deriva de una curiosidad natural acerca del mundo. La mayoría de la gente tiene su primera experiencia de investigación cuando su curiosidad los lleva a plantearse: "me pregunto qué pasaría si...", o "me pregunto por qué...", seguido por un intento de responder dicha pregunta. ¿Cuáles son las fuentes de inspiración para preguntas como éstas? ¿Cómo formularlas a partir de las ideas de otras personas y de investigación previa? En este capítulo exploraremos algunas fuentes de ideas científicas. También consideraremos la naturaleza de los reportes de investigación que se publican en revistas especializadas.

HIPÓTESIS Y PREDICCIONES

En su mayoría, los estudios de investigación intentan probar una **hipótesis** formulada por el investigador. Una hipótesis es, en realidad, un tipo de idea o pregunta; plantea una afirmación sobre algo que puede ser verdadero. Por lo tanto, una hipótesis es únicamente una idea o pregunta tentativa que espera evidencias que la sustenten o refuten. En ocasiones, las hipótesis son preguntas muy generales e informales. Por ejemplo, Geller, Russ y Altomari (1986) tenían preguntas generales sobre el consumo de cerveza de los estudiantes universitarios: ¿cuántas cervezas ingieren en una visita promedio a un bar universitario? ¿Cuánto tiempo permanecen ahí los estudiantes? ¿Existen diferencias en el consumo entre mujeres y hombres? Con este tipo de preguntas en mente, los investigadores desarrollaron un procedimiento para recolectar datos y responder las preguntas. Los problemas de investigación formulados de esta manera constituyen hipótesis informales o simples preguntas sobre el comportamiento.

Con frecuencia, las hipótesis se formulan en términos más específicos y formales. Generalmente, las hipótesis formales establecen que dos o más variables están relacionadas entre sí. Así, los investigadores pueden formular hipótesis como: "el hacinamiento da como resultado un desempeño pobre en tareas cognoscitivas" o "poner atención a más características de algo que se va a aprender, da como resultado una memoria más amplia". Hipótesis como éstas se formulan con base en hallazgos de investigaciones pasadas y consideraciones teóricas. Entonces, el investigador diseña un estudio para someter a prueba la hipótesis. En el ejemplo sobre el hacinamiento, el investigador podría colocar a los participantes en una habitación abarrotada o en un cuarto con poca gente, para trabajar en una serie de tareas y en esas condiciones observar su desempeño.

En este punto, el científico haría una **predicción** específica con respecto al resultado del experimento. Aquí, el pronóstico podría ser que "los participantes en la condición con poca gente tendrán un mejor desempeño en las tareas que los participantes en la condición de hacinamiento". Si esta predicción se confirma con los resultados del estudio, se sustenta la hipótesis. Si la predicción no se confirma, el investigador podrá rechazar la hipótesis (y creerá que el hacinamiento no produce un desempeño pobre) o realizará más investigaciones con el empleo de diferentes métodos para estudiar la hipótesis. Es importante señalar

que cuando los resultados de un estudio confirman una predicción, únicamente se *sustenta* la hipótesis, no *se prueba*. Los investigadores estudian la misma hipótesis con una diversidad de métodos, y cada vez que esta hipótesis es apoyada por un estudio de investigación, se incrementa nuestra confianza en que es correcta.

A QUIÉN ESTUDIAMOS: UNA NOTA SOBRE LOS TÉRMINOS

Hemos estado utilizando el término *participantes* para referirnos a los individuos que colaboran en proyectos de investigación. Un término equivalente en la investigación psicológica es el de *sujetos.* El *Manual de Publicaciones de la Asociación Psicológica Americana* (American Psychological Association, 2001) recomienda el uso del término *participantes* cuando se describe a seres humanos que toman parte en investigaciones psicológicas. Usted encontrará ambos términos cuando lea investigaciones, y en este libro utilizaremos los dos. Entre otros términos que hallará se incluyen *encuestado* e *informante.* Los individuos que toman parte en investigaciones con encuestas suelen denominarse *encuestados.* Los *informantes* son aquellos que ayudan a los investigadores a comprender la dinámica de ambientes culturales y organizacionales particulares; este término se originó en las investigaciones antropológicas y sociológicas, y en la actualidad también es utilizado por los psicólogos.

FUENTES DE IDEAS

No es fácil determinar de dónde provienen las buenas ideas. Muchas personas son capaces de tener ideas muy valiosas, pero les es difícil expresar verbalmente el proceso por medio del cual se generaron. Los caricaturistas saben esto: representan una buena idea como un foco brillante sobre la cabeza de un personaje; pero, ¿de dónde proviene la electricidad? Consideremos cinco fuentes de ideas: el sentido común, la observación del mundo que nos rodea, las teorías, investigaciones anteriores y los problemas prácticos.

El sentido común

Una fuente de ideas que pueden ser probadas es el cuerpo de conocimientos llamado sentido común, las cosas que todos creemos que son verdaderas. "¿Los opuestos se atraen" o "¿Dios los hace y ellos se juntan?". Si usted "recurre al castigo", entonces "¿echa a perder al niño?", "¿una imagen vale más que mil palabras?". Plantearse preguntas como éstas puede conducirnos a elaborar programas de investigación que estudien la atracción, los efectos del castigo y el papel de las imágenes visuales en el aprendizaje y la memoria.

Someter a prueba una idea del sentido común puede ser valioso, ya que ideas como éstas no siempre resultan correctas, o las investigaciones pueden

demostrar que el mundo real es mucho más complicado de lo que nuestras ideas del sentido común lo hacen parecer. Por ejemplo, bajo ciertas circunstancias las imágenes pueden mejorar la memoria, pero en ocasiones las imágenes desvían el aprendizaje (véase Levin, 1983). Llevar a cabo investigaciones para probar ideas del sentido común a menudo nos obliga a ir más allá de una teoría de sentido común sobre el comportamiento.

La observación del mundo que nos rodea

Las observaciones de acontecimientos personales y sociales pueden proporcionar muchas ideas de investigación. La curiosidad que le despiertan sus observaciones y experiencias puede llevarlo a plantearse preguntas sobre todo tipo de fenómenos. De hecho, esta clase de curiosidad es la que motiva a muchos estudiantes a involucrarse en su primer proyecto de investigación.

¿Alguna vez ha tenido la experiencia de guardar algo en un "lugar especial", con la certeza de encontrarlo después (y donde nadie más podría buscarlo), para descubrir más tarde que no puede recordar dónde lo puso? Una experiencia como ésta podría producir una investigación sistemática sobre sí es una buena idea guardar cosas en lugares especiales. De hecho, Winograd y Soloway (1986) realizaron una serie de experimentos sobre este tema. Su investigación demostró que las personas tienden a olvidar dónde guardaron algo cuando se presentan dos condiciones: 1) el lugar donde se coloca el objeto se considera muy fácil de recordar, *y 2)* la ubicación se considera un lugar muy improbable para el objeto. Por lo tanto, aunque parezca una buena idea en ese momento, guardar algo en un lugar poco común suele ser una mala idea.

Un ejemplo más reciente demuestra la diversidad de ideas que pueden generarse, debido a la curiosidad con respecto de las cosas que suceden a nuestro alrededor. Durante los últimos años, ha habido una gran controversia sobre los efectos de la letra de las canciones, con el temor de que cierto tipo de música rock y rap produzca promiscuidad sexual, consumo de drogas y violencia. Algunos grupos, como el Centro de recursos musicales de padres (Parents' Music Resource Center, PMRC) , desean censurar las letras de ciertas canciones y han persuadido a compañías disqueras para que coloquen etiquetas de advertencia en los discos. Se han realizado audiencias en el Congreso sobre este tema. Algunos científicos han decidido llevar a cabo investigaciones para examinar temas que han surgido por esta controversia. Fried (1999) sugirió que la reacción negativa, ante la música rap en particular, pudo haber surgido debido a su asociación con la música negra. Para probar esta idea, Fried pidió a algunos participantes que leyeran la letra de una canción popular llamada "Bad Man's Blunder", que tiene un tema violento; se les dijo que la canción era música rap o campirana *(country).* Aun cuando es claro que no se trata de un rap, los participantes mostraron una reacción más negativa ante la letra cuando se les dijo que se trataba de música rap, que cuando se les presentó como música campirana.

El mundo que nos rodea es una rica fuente de material para investigación científica. Cuando el psicólogo Robert Levine era profesor en Brasil, hace varios

años, notó que los estudiantes brasileños eran más informales para llegar a clases a tiempo y para irse que sus contrapartes en Estados Unidos. Esa observación lo condujo a iniciar estudios sobre el ritmo de vida en diversos países, así como en varias ciudades de Estados Unidos (Levine, 1990). Investigaciones como éstas ilustran un punto que se comentó en el capítulo 1: las opiniones y experiencias personales proporcionan comprensión sobre el comportamiento, pero el pensamiento científico requiere que probemos nuestras ideas por medio de la investigación.

Finalmente, debemos mencionar el papel del *serendipity*, en ocasiones los descubrimientos más interesantes son el resultado de un accidente o de la buena suerte. Ivan Pavlov es mejor conocido por el descubrimiento del llamado condicionamiento clásico; si un estímulo neutral (como un tono) se aparea de forma repetida con un estímulo incondicionado (comida) que produce una respuesta refleja (salivación), eventualmente producirá la respuesta cuando se presente solo. Pavlov no esperaba descubrir el condicionamiento clásico, sino que estaba estudiando el sistema digestivo en perros, para lo cual medía su salivación cuando los alimentaba. Él descubrió accidentalmente que los perros salivaban antes de ser alimentados, y entonces estudió la forma en que los estímulos que precedían a la alimentación podrían producir una respuesta de salivación. Por supuesto, este tipo de descubrimientos accidentales ocurre únicamente cuando se observa el mundo con un ojo curioso.

Teorías

Gran cantidad de investigaciones en las ciencias del comportamiento prueba teorías de la conducta. Las teorías ofrecen dos funciones importantes para una mayor comprensión del comportamiento. Primero, las teorías *organizan y explican* una diversidad de hechos o descripciones específicas de la conducta. Esos hechos y descripciones no tienen mucho significado en sí mismos, por lo cual se requiere de las teorías para encuadrarlos en un marco de referencia que permite que el mundo sea más comprensible al proporcionar unos cuantos conceptos abstractos, con los cuales podemos organizar y explicar una gran diversidad de conductas. Por ejemplo, considere cómo la teoría de la evolución de Charles Darwin organizó y explicó una variedad de sucesos con respecto de las características de las especies animales. De manera similar, en la psicología una teoría sobre la memoria asevera que existen sistemas separados de memoria a corto plazo y memoria a largo plazo. Esta teoría explica ciertas observaciones específicas sobre el aprendizaje y la memoria, e incluye además fenómenos como los distintos tipos de déficits de memoria que resultan de un golpe en la cabeza, en oposición al daño del área del hipocampo cerebral, y la proporción en que la persona olvida información que acaba de leer.

En segundo lugar, las teorías *generan nuevos conocimientos* cuando enfocan nuestro pensamiento de tal manera que notamos nuevos aspectos de la conducta; las teorías guían nuestras observaciones del mundo. Las teorías generan hipótesis sobre el comportamiento y el investigador realiza estudios para veri-

ficar si son correctas. *Si* los estudios confirman las hipótesis, entonces la teoría se sustenta. Conforme se acumula mayor evidencia consistente con la teoría, crece nuestra confianza en que ésta es correcta. Sin embargo, las investigaciones también pueden revelar debilidades en una teoría y obligar a los especialistas a modificarla o a crear una nueva, más detallada e incluyente.

La teoría de la evolución continúa ayudando a los psicólogos a generar hipótesis (Buss, 1998; Gaulin y McBurney, 2000). Por ejemplo, esta teoría afirma que los hombres y las mujeres han desarrollado distintas estrategias de reproducción. Todos los individuos tienen un interés evolutivo en transmitir sus genes a las futuras generaciones. No obstante, las mujeres tienen relativamente pocas oportunidades de reproducirse; cuentan con un rango de edad limitado para hacerlo y deben dedicar una enorme cantidad de tiempo y energía al cuidado de sus hijos. Los hombres, en contraste, se pueden reproducir en cualquier momento y tienen una ventaja reproductiva al engendrar tantos descendientes como les sea posible. Debido a estas diferencias, la teoría predice que las mujeres y los hombres utilizarán distintos criterios para seleccionar una pareja. Las mujeres se interesarán más en hombres que puedan ofrecer apoyo para la crianza de los hijos, aquellos con un estatus más alto, con más recursos económicos y dominantes. Los hombres elegirán mujeres más jóvenes, más saludables y más atractivas físicamente. Las investigaciones en diversas culturas confirman estas predicciones (Buss, 1989). Aun cuando los estudios apoyan la teoría de la evolución, se pueden crear teorías alternativas que expliquen mejor los mismos hallazgos; las teorías tienen vida y son dinámicas. Eagly y Wood (1999) interpretaron la investigación de Buss en términos de la estructura social. Argumentan que las variaciones en resultados por género surgen del hecho de que existen diferencias entre hombres y mujeres en la división del trabajo —los varones son responsables del bienestar económico de la familia y las mujeres, del cuidado de los hijos-- y estas disimilitudes explican las discrepancias de género en las preferencias al elegir pareja. Los estudios sobre este interesante tema continúan.

Las teorías se modifican conforme nuevas investigaciones limitan su ámbito. La necesidad de transformar las teorías se ejemplifica con la teoría de la memoria a corto y largo plazos, mencionada anteriormente. La propuesta original del sistema de memoria a largo plazo la describía como un almacén de recuerdos fijos y permanentes. Sin embargo, investigaciones realizadas por psicólogos cognoscitivos, entre ellos Loftus (1979), han demostrado que los recuerdos se reconstruyen y reinterpretan fácilmente. En un estudio, los participantes observaron una película sobre un accidente automovilístico, y después se les pidió que relataran lo que vieron. Loftus encontró que los recuerdos de los participantes estaban influidos por la forma en que eran cuestionados. Por ejemplo, a los participantes que les preguntó si habían observado "el" faro descompuesto, presentaron una mayor tendencia a responder que sí en comparación con los participantes a quienes les preguntó si habían observado "un" faro descompuesto. Resultados como éstos requieren una teoría más compleja sobre la forma en que opera la memoria a largo plazo.

Investigaciones previas

Una cuarta fuente de ideas es la investigación previa. La familiarización con el cuerpo de investigación sobre un tema, tal vez constituya la mejor forma de generar nuevas ideas de investigación. Puesto que los resultados de los estudios se publican, los investigadores pueden utilizar el material de la literatura previa sobre un tema para refinar y expandir de forma continua nuestros conocimientos. Virtualmente, cada estudio hace surgir preguntas que pueden estudiarse en investigaciones posteriores. La investigación puede provocar intentos de aplicar los hallazgos en un ambiente diferente, de estudiar el tema con un grupo de edad distinto o de utilizar una metodología diferente para replicar los resultados. El estudio de Geller, *et al.* (1986), sobre el consumo de cerveza reportó que los estudiantes que compran bebidas por vaso consumen más cerveza que aquellos que compran botellas; que los hombres beben más que las mujeres; que los estudiantes en grupo beben más que quienes están solos, y que las mujeres permanecen más tiempo en el bar que los hombres. Los conocimientos derivados de este estudio y sus resultados podrían inspirar, por ejemplo, investigaciones sobre formas de reducir el consumo excesivo de bebidas alcohólicas entre los estudiantes universitarios.

Además, conforme usted se familiariza más con la literatura de investigación sobre un tema, puede observar inconsistencias en los resultados de los estudios que deben ser investigados; o tal vez desee estudiar explicaciones alternativas para los resultados. Por otro lado, lo que usted conoce sobre un área de investigación, con frecuencia puede aplicarse con éxito en otro campo de estudio.

Veamos un ejemplo concreto de un estudio que fue diseñado para señalar las fallas metodológicas de investigaciones previas. El estudio se refería a un método para ayudar a niños diagnosticados como autistas. El autismo infantil se caracteriza por un número de síntomas que incluyen un deterioro grave en el lenguaje y las habilidades de comunicación. Recientemente, los padres y el personal de salud se han interesado en una técnica llamada "comunicación facilitada", que aparentemente permite que un niño autista se comunique con los demás cuando presiona letras y otros símbolos de un teclado. Un facilitador sostiene la mano del niño para permitir que la habilidad de éste determine qué tecla va a presionar. Con esta técnica, una cantidad de niños autistas comienzan a comunicar sus pensamientos y sentimientos, y a responder preguntas que se les plantean. La mayoría de las personas que observan la "comunicación facilitada" en acción considera que la técnica es un avance milagroso.

La conclusión de que la "comunicación facilitada" es eficaz, se basa en la comparación de las habilidades de los niños autistas para comunicarse con y sin el facilitador. La diferencia es impresionante para la mayoría de los observadores. Sin embargo, recuerde que los científicos son escépticos por naturaleza; examinan todas las evidencias cuidadosamente y se preguntan si las aseveraciones están justificadas. En el caso de la "comunicación facilitada", Montee, Miltenberger y Wittrock (1995) señalaron que el facilitador puede estar guiando sin intención los dedos del niño para escribir oraciones con significado. En

otras palabras, es el facilitador, y no el individuo autista, quien controla la comunicación. Montee y sus colaboradores llevaron a cabo un estudio para probar esta idea. En una condición, se les mostró una fotografía tanto al facilitador como al niño autista; a este último se le pidió que indicara lo que aparecía en la fotografía y escribiera la respuesta con el facilitador. Esto se hizo en varios ensayos. En otra condición, únicamente el niño veía la fotografía. En una tercera condición, se les mostraron fotografías diferentes al niño y al facilitador (aunque el facilitador no sabía esto). En concordancia con la hipótesis de que el facilitador controla las respuestas del niño, las fotografías fueron identificadas correctamente sólo en la condición en que ambos veían la misma fotografía. Además, cuando el niño y el facilitador observaron fotografías diferentes, el infante nunca dio la respuesta correcta, y con frecuencia identificaba la fotografía que había visto el facilitador.

Problemas prácticos

La investigación también se ve estimulada por problemas prácticos en los que puede tener aplicaciones inmediatas. Por ejemplo, grupos de urbanistas y ciudadanos podrían consultar a los ciclistas para determinar la ruta más deseable para una ciclopista. A una mayor escala, los especialistas han guiado la política pública cuando realizan investigaciones sobre los efectos de la exposición a materiales pornográficos, así como otros temas sociales y de salud. Gran parte de la investigación aplicada y evaluativa descrita en el capítulo 1 abarca temas como éstos.

INVESTIGACIÓN BIBLIOGRÁFICA

Antes de realizar cualquier proyecto de investigación, un científico debe conocer con profundidad los hallazgos de estudios previos. Aun cuando la idea básica ya haya sido formulada, la revisión de escritos anteriores ayudará a que el investigador aclare la idea y diseñe el estudio. Por consiguiente, es importante saber cómo buscar la literatura sobre un tema y cómo leer los reportes de investigación en las revistas especializadas. En esta *sección* hablaremos únicamente de los fundamentos de la investigación bibliográfica, para mayor información, puede remitirse a una guía más detallada sobre investigación bibliográfica en psicología y elaboración de artículos de revisión de investigaciones, como las de Reed y Baxter (2003) y Rosnow y Rosnow (2002).

La naturaleza de las revistas científicas

Si usted ha recorrido la sección de revistas de su biblioteca, habrá notado la gran cantidad de publicaciones científicas. En éstas los investigadores reportan los resultados de sus estudios. Una vez que un proyecto de investigación se ha com-

pletado, el estudio se escribe en un reporte, el cual después puede enviarse al editor de una revista apropiada. El editor solicita la revisión de otros científicos en el mismo campo, y después decide si se acepta el reporte para su publicación. Dado que cada revista tiene un espacio limitado y recibe más artículos de los que puede publicar, la mayoría de éstos son rechazados. Los que son aceptados, se publican aproximadamente un año después.

La mayoría de las revistas de psicología se especializan en una o dos áreas del comportamiento humano o animal. Aun así, el número de revistas en muchas áreas es tan grande que es casi imposible que una persona pueda leer todas. La tabla 2.1 muestra algunas de las publicaciones más importantes en diversas áreas de la psicología. La tabla no incluye revistas que se editan en Internet; tampoco toma en cuenta muchas revistas que se publican en áreas que están estrechamente relacionadas con la psicología, ni áreas muy especializadas dentro de la psicología. Claramente sería difícil leer todas las revistas de la lista, incluso si restringiera su lectura a una sola rama de investigación en psicología, como el aprendizaje y la memoria. Si busca investigaciones sobre un tema específico, no sería práctico buscar en cada número de cada revista en la que podrían publicarse investigaciones relevantes. Por fortuna, usted no necesita hacer esto.

Psychological Abstracts *(resúmenes psicológicos)*

La Asociación Psicológica Americana comenzó la publicación de los *Psychological Abstracts*, en 1927. Hasta hace poco tiempo, los estudiantes realizaban búsquedas de literatura manualmente, con el fin de encontrar los resúmenes breves *(abstracts)* de artículos sobre psicología y disciplinas relacionadas que se publicaban cada mes en los *Psychological Abstracts;* aún se edita su versión impresa. No obstante, en la actualidad es más probable que usted realice búsquedas utilizando bases de datos por computadora que contengan los resúmenes. El sistema de las bases de datos por computadora de la Asociación Psicológica Americana se denomina *PsycINFO;* es muy fácil acceder a ella por medio de Internet, y se actualiza cada mes. La cobertura de *PsycINFO* comenzó en 1966, aunque se puede encontrar información de años anteriores. Una base de datos relacionada y que se utiliza en algunas bibliotecas es el *PsycFIRST*. Éste contiene resúmenes de los tres años anteriores. Los procedimientos exactos para realizar una búsqueda en los sistemas del *PsycINFO* dependen de los arreglos que su biblioteca haya hecho para tener acceso a la base de datos. En todos los casos, obtendrá una lista de resúmenes que están relacionados con su tema de particular interés. Después puede encontrar y leer los artículos en la biblioteca. Si un manuscrito importante no está disponible ahí, pregunte al bibliotecario sobre los servicios para obtener textos de otras bibliotecas.

Cómo hacer una búsqueda en PsycINFO

La forma exacta del sistema que utilizará para buscar en el *PsycINFO dependerá del sistema de su computadora.* La figura 2.1 (página 28) ilustra un sistema basado

TABLA 2.1 Algunas de las principales revistas de psicología

General

*American Psychologist** (artículos generales sobre diversos temas)
*Contemporary Psychology** (revisiones de libros)
*Psychological Bulletin** (revisiones de la literatura)
*Psychological Review** (artículos teóricos)
Psychological Science
*Psychological Methods**
Current Directions in Pychological Science
Psychological Science in the Public Interest

Áreas experimentales de la psicología

*Journal of Experimental Psychology: General**
*Journal of Experimental Psychology: Applied**
*Journal of Experimental Psychology: Learning Memory and Cognition**
*Journal of Experimental Psychology: Human Perception and Performance**
*Journal of Experimental Psychology: Animal Behavior Processes**
*Journal of Comparative Psychology**
*Behavioral Neuroscience**
Bulletin of the Psychonomic Society
Learning and Motivation
Memory and Cognition
Cognitive Psychology
Cognition
Cognitive Science
Discourse Processes
Journal o f the Experimental Analysis of Behavior
Animal Learning and Behavior
*Neuropsychology **
*Emotion**
*Experimental and Clinical Psychopharmacology**

Psicología clínica y consejería

*Journal of Abnormal Psychology**
*Journal of Consulting and Clinical Psychology**
*Journal of Counseling Psychology**
Behavior Research and Therapy
Journal of Clinical Psychology
Behavior Therapy
Journal of Abnormal Child Psychology
Journal of Social and Clinical Psychology
*Professional Psychology: Research and Practice**

(continúa)

TABLA 2.1 Algunas de las principales revistas de psicología *(continuación)*

Psicología del desarrollo

*Developmental Psychology**
*Psychology and Aging**
Child Development
Journal of Experimental Child Development
Journal of Applied Developmental Psychology
Developmental Review
Infant Behavior and Development
Experimental Aging Research
Merril-Palmer Quarterly

Psicología social y de la personalidad

*Journal of Personality and Social Psychology**
Personality and Social Psychology Bulletin Journal of Experimental Social Psychology Journal of Research in Personality Journal of Social Issues Social Psychology Quarterly Journal of Applied Social Psychology Basic and Applied Social Psychology Journal of Social and Personal Relationships

Áreas de la psicología aplicada

*Journal of Applied Psychology **
*Journal of Educational Psychology**
Journal of Applied Behavior Analysis
*Health Psychology**
*Psychological Assessment**
*Psychology, Public Policy and Late**
Law and Human Behavior
Educational and Psychological Measurement
American Educational Research Journal
Evaluation Review
Evaluation and Program Planning
Environment and Behavior
Journal of Environmental Psychology
Journal of Consumer Research
Journal of Marketing Research

TABLA 2.1 Algunas de las principales revistas de psicología *(continuación)*

Estudios de la familia y conducta sexual

Journal of Family Psychology Journal of Marriage and the Family Journal of Marital and Family Therapy Journal of Sex Research Journal of Sexual Behavior Journal of Homosexuality*

Temas étnicos, de género y transculturales

Hispanic Journal of Behavioral Sciences
Journal of Black Psychology
Sex Roles
Psychology of Women Quarterly
Journal of Cross-Cultural Psychology
*Cultural Diversity and Ethnic Minority Psychology**

Algunas revistas canadienses y británicas

Canadian Journal of Experimental Psychology
Canadian Journal of Behavioral Science British Journal of Psychology British Journal of Social and Clinical Psychology

* Publicados por la Asociación Psicológica Americana.

en Internet (WebSPIRS de SilverPlatter); la misma base de datos de información del *PsycINFO* se obtiene de diferentes maneras por medio de otros sistemas. Esta pantalla muestra las tres partes principales de una búsqueda: debe escribir el tema que desea, debe ser capaz de ver un despliegue de los resultados de la búsqueda y tener opciones de ayuda para su propósito.

Su tarea más importante es especificar los términos de la búsqueda que desea que la computadora utilice. Estos se anotan en un recuadro de entrada (en la figura 2.1 se le denomina *Find box*, "recuadro para encontrar"). ¿Cómo saber qué palabras se deben escribir en el recuadro de entrada? Lo más común es el uso de términos psicológicos estándar. El *Thesaurus of Psychological Index Terms* contiene todos los vocablos estándar que se emplean para ordenar los resúmenes, y pueden obtenerse directamente en la mayoría de los sistemas del *PsycINFO*. Suponga que usted está interesado en el tema de la ansiedad en los exámenes. Resulta que *examen y ansiedad* son descriptores importantes en el diccionario, o *thesaurus*. Si busca bajo *ansiedad*, aparecerán todos los términos relacionados,

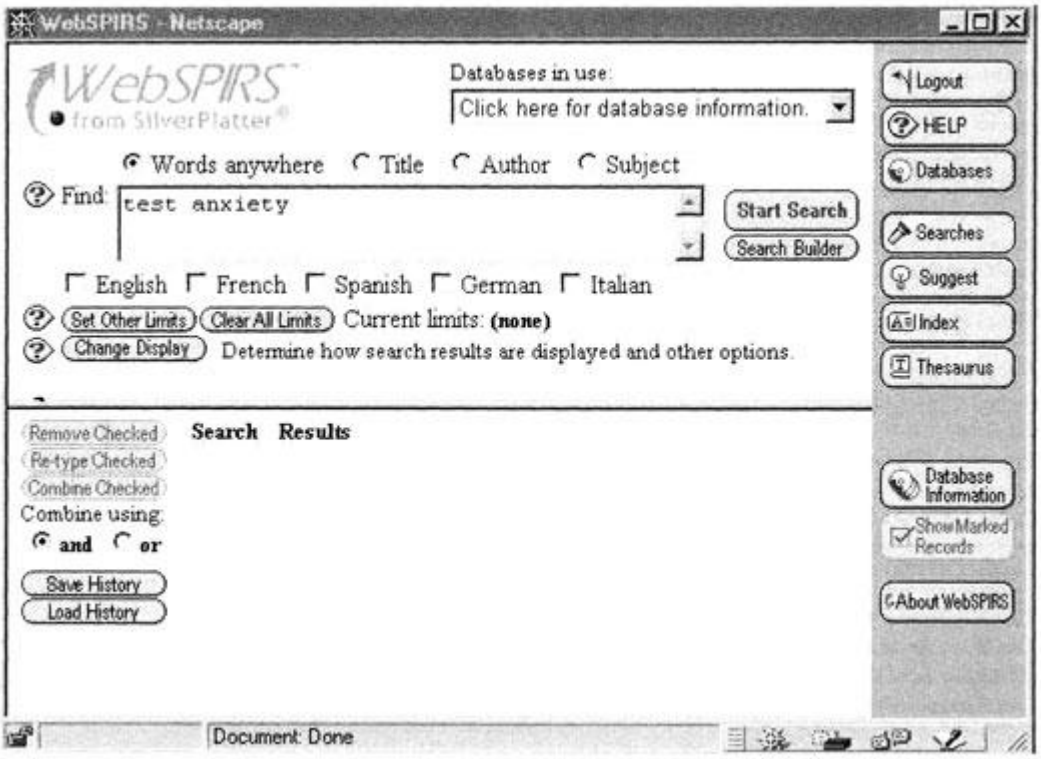


FIGURA 2.1 Pantalla del usuario de *PsycINFO.*

Ventana del navegador Netscape © 2002 Netscape Communications Corporation. Utilizado bajo permiso. Netscape Communications no ha autorizado, financiado, endosado ni aprobado esta publicación y no es responsable de su contenido. Reproducido bajo permiso de la Asociación Psicológica Americana, editora de la base de datos *PsycINFO,* todos los derechos reservados. No puede ser reproducido sin permiso.

como *ansiedad por separación, ansiedad social* y *ansiedad en los exámenes.* Si utiliza el diccionario, puede verificar cualquier término y después pedir una búsqueda de éste. No obstante, supongamos que usted está utilizando una ventana de búsqueda estándar, como en la figura 2.1. Cuando dé la orden para el inicio de la búsqueda, aparecerán los resultados.

Aquí se presentan los resultados de uno de los artículos encontrados en una búsqueda sobre la ansiedad ante los exámenes. Se eligió la aparición de una gran cantidad de información sobre el artículo, para ejemplificar la manera en que está organizada la información en la base de datos; generalmente, usted elegirá la aparición de menos datos. Observe que el resultado está organizado en "campos" de información. Aquí se incluye el nombre completo de cada campo, aunque muchos sistemas manejan abreviaturas. Casi siempre, usted preferirá ver el

título (abreviado como TI), el autor (AU), la fuente (SO) y el resumen o "abstract" (AB). Observe que también aparecen campos como el tipo de documento, una frase clave que describe brevemente el artículo, descriptores principales y secundarios, y el grupo de edad.

TIPO DE DOCUMENTO Artículo de revista científica TÍTULO: Individual differences in students' retention of knowledge and conceptual structures learned in university and high school courses: the case of test anxiety.

AUTOR: Naveh-Benjamin, -Moshe; Lavi, -Hagit; McKeachie, -Wilbert-J.; Lin, -Yi-Guang

AFILIACIÓN DEL PRIMER AUTOR: Ben-Gurion U of the Negev, Dept. de Ciencias del comportamiento, Beer-Sheva, Israel

FUENTE: Applied-Cognitive-Psychology. 1997 Dec; Vol 11(6): 507-526

ISSN: 0888-4080

AÑO DE PUBLICACIÓN: 1997

RESUMEN: Fueron examinadas las diferencias individuales en la retención de conocimientos de los estudiantes, varios años después de estudiar material académico. La evaluación de la retención de materiales, en función de la ansiedad de los estudiantes ante los exámenes, puede permitir evaluar si el deficiente desempeño académico y la organización del material por parte de los alumnos con altos niveles de ansiedad ante los exámenes se debe a una insuficiencia en la memoria o a una falla en la organización del aprendizaje y de los conocimientos. En dos estudios, 210 estudiantes estadounidenses (estudio 1) y 258 israelíes (de entre 17 y 27 años, estudio 2), con diferentes niveles de ansiedad ante los exámenes, completaron tareas que nos permitieron evaluar sus niveles de conocimientos y su organización cognoscitiva del material. Las tareas se administraron al final del curso o a diferentes intervalos de retención, hasta siete años después de finalizar el aprendizaje. Los resultados indicaron que mientras los estudiantes muy ansiosos ante los exámenes presentaron peores puntuaciones

al final de los cursos, con relación a otros estudiantes, en pruebas de conocimientos y de organización cognoscitiva, los alumnos muy ansiosos ante las pruebas, examinados en diversos intervalos de retención después de dos cursos, se desempeñaron tan bien como los demás estudiantes. Se discuten las implicaciones teóricas y prácticas de estos resultados. (© 1998 APA/*PsycINFO*, todos los derechos reservados).

FRASE CLAVE: Ansiedad ante las pruebas y diferencias individuales en la retención del conocimiento y estructuras conceptuales aprendidas en cursos universitarios y en la preparatoria, estudiantes estadounidenses e israelíes.

DESCRIPTORES PRINCIPALES: *diferencias-individuales, *retención-; *ansiedad-exámenes.

DESCRIPTORES SECUNDARIOS: Adultez-; estudiantes-universitarios ; estudiantes-preparatoria.

GRUPO DE EDAD: Adultos.

POBLACIÓN: humanos, hombres, mujeres.

UBICACIÓN DE LA POBLACIÓN: Israel; Estados Unidos.

TIPO DE PUBLICACIÓN: Estudio-empírico.

Cuando realiza una búsqueda simple con una sola palabra o frase, como *ansiedad ante exámenes,* la búsqueda muestra artículos que contienen la palabra o frase en cualquiera de los campos listados. Con frecuencia encontrará que esto resulta en demasiados artículos, incluso algunos que no son directamente relevantes para sus intereses. Una forma de acortar la búsqueda es limitarla a ciertos campos. Por ejemplo, usted podría especificar *ansiedad ante exámenes en TI* para limitar su búsqueda a los artículos que contienen este término en el título. De forma similar, usted puede especificar que la búsqueda sólo dé resultados de artículos científicos (ni libros ni tesis) o que incluya participantes de ciertos grupos de edades.

El uso de operadores boleanos como *AND, OR y NOT* también le pueden ayudar en su búsqueda. Suponga que desea restringir la búsqueda de *ansiedad ante exámenes en TÍTULO,* únicamente en estudiantes universitarios. Puede hacer esto si pide *(ansiedad ante exámenes en TITULO) AND (estudiantes universitarios).* La palabra AND obliga a que se cumplan ambas condiciones para incluir un artículo. Los paréntesis se emplean para separar las distintas partes de la es-

pecificación de su búsqueda, y son útiles cuando su búsqueda se vuelve más complicada. De hecho, podrían haberse dejado fuera de esta exploración, pero se incluyen como un ejemplo.

La operación OR se utiliza para ampliar una búsqueda demasiado estrecha. Suponga que desea encontrar artículos que hablan sobre las relaciones amorosas en Internet. Buscamos únicamente *Internet AND romance y* encontramos 15 artículos; al cambiar las especificaciones a *Internet AND (romance OR citas OR amor OR atracción),* resultaron 51 artículos. Los artículos que incluyen el término *Internet y cualquiera* de los otros vocablos especificados, fueron incluidos en la búsqueda.

La operación NOT excluye resúmenes basados en un criterio que usted especifica. La operación NOT se utiliza cuando usted anticipa que los criterios de exploración serán satisfechos por algunos resúmenes irrelevantes. En el ejemplo de Internet, es posible que la palabra *atracción* encuentre investigaciones sobre entrevistas de empleo por Internet, en las que se describa la atracción por un empleo o por un candidato particular a un trabajo. Para excluir el término *empleo* de los resultados de la búsqueda, se pueden hacer los siguientes ajustes: *Internet AND (romance OR citas OR amor OR atracción) NOT empleo.* Cuando se *realizó* esta búsqueda, encontré 47 resúmenes, en lugar de los 51 obtenidos previamente.

Otra útil herramienta de búsqueda es el "comodín" asterisco (*). El asterisco implica cualquier conjunto de letras en una palabra, por lo cual puede expandir su búsqueda. Considere la palabra *romance* en la exploración anterior; al emplear *román**, la búsqueda se ampliará e incluirá *romance y romántico.* El comodín puede ser muy útil con términos como *infan** para encontrar *infante, infancia, infantil,* etc. Sin embargo, debe ser cuidadoso con esto; ¡la búsqueda *román** también resultaría en *romanos y romanticismo!* En este caso, podría ser más eficiente si sencillamente se añade *OR romántico* a la búsqueda. Estas estrategias de búsqueda se resumen en la tabla 2.2.

TABLA 2.2 Algunas estrategias de búsqueda en *PsycINFO*

Estrategia 1: utilice campos como TI y AU.

Ejemplo: *(divorcio) en TI,* para pedir que un término aparezca en el título. Estrategia 2: utilice AND para limitar la búsqueda.

Ejemplo: *divorcio AND niño,* se requieren ambos términos.

Estrategia 3: utilice OR para expandir la búsqueda.

Ejemplo: *divorcio OR rompimiento,* incluye ambos términos. Estrategia 4: utilice NOT para excluir palabras en la búsqueda.

Ejemplo: *timidez NOT terapia,* excluye cualquier artículo sobre timidez que incluya el término *terapia.* Estrategia 5: utilice el comodín asterisco (*).

Ejemplo: *infan** encuentra cualquier palabra que se inicie con *infan* (infancia, infante, etcétera).

Cuando termine su búsqueda, podrá imprimir los resultados. Cuando imprima, podrá elegir cuál de los campos desea que aparezcan. Probablemente usted no necesite todos los campos presentados en el ejemplo anterior. Muchos investigadores prefieren guardar los resultados de la búsqueda en un disquete. La información puede utilizarse después con otros programas, como un procesador de texto o un ordenador de citas. Si no tiene acceso a un disco, también puede enviar los resultados de la búsqueda a su dirección de correo electrónico.

Science Citation Index *y* Social Science Citation Index

Dos recursos de búsqueda son el *Science Citation Index* (SCI) y el *Social Science Citation Index* (SSCI). Ambos le permiten buscar a través de información de citas, como el nombre del autor o el título del artículo. El SCI incluye disciplinas como la biología, química, biomedicina y farmacología, mientras que el SSCI comprende ciencias sociales y del comportamiento, como la sociología y la justicia criminal. La característica más importante de ambos recursos es su capacidad para utilizar el método del "artículo clave". Usted primero necesita identificar un "artículo clave" en su tema, por lo general, uno que fue publicado anteriormente y que es especialmente relevante a sus intereses. Después, puede buscar artículos subsecuentes que citan al artículo clave. Esta búsqueda le dará una bibliografía de artículos relacionados con su tema. Para ofrecer un ejemplo de este proceso, elegiremos el siguiente artículo:

> Wells, G. L. y Bradfield, A. L. (1999). Distortions on eyewitnesses' recollections: Can the postidentification feedback effect be moderated? *Psychological Science, 10,* 138-144.

Cuando realizamos una búsqueda de artículos utilizando el SSCI, encontramos diez artículos que habían citado el trabajo de Wells y Bradfield, desde que fue publicado en 1999. He aquí uno de ellos:

> Ghetti, S., Qin, J. J. y Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective expenence. *Developmental Psychology, 38,* 705-718.

El siguiente paso es familiarizarse con este artículo, al igual que con los demás de la lista. Puede resultar que uno o más de los artículos se conviertan en "artículos clave" para nuevas búsquedas.

También se puede especificar una "persona clave" para encontrar todos los artículos escritos por un autor, o se cita a una persona en particular a partir de cierta fecha. Las búsquedas en el SCI y en el SSCI son mucho más fáciles si se realizan en una computadora; están disponibles en CD-ROM o en Internet.

Revisiones de la literatura

También son útiles los artículos que resumen las investigaciones en un área en particular. El *Psychological Bulletin* publica revisiones de la literatura en diversos temas de psicología. Cada año, el *Annual Review of Psychology* publica artículos que resumen los desarrollos recientes en varias áreas de la psicología. Otras disciplinas tienen revisiones anuales similares.

Otros recursos electrónicos de búsqueda

Su biblioteca puede tener o no acceso al *PsycINFO* o a las bases de datos SCI y SSCI. El número de bases de información que una biblioteca puede comprar en la actualidad es enorme; el presupuesto y otras consideraciones determinan cuáles están disponibles. Usted necesitará aprovechar los materiales didácticos que su biblioteca ofrece para ayudarle a aprender una mejor forma de buscar la información de que dispone. Otras bases de datos importantes incluyen First-Search, Sociological Abstracts, MEDLINE y ERIC *(Educational Resources Information Center).* Además, servicios como el Lexis-Nexis y Factiva le permiten buscar recursos generales de medios de comunicación, como periódicos. Un bibliotecario puede ayudarle a utilizar éstos y otros recursos que están disponibles para usted.

Algunos de los recursos de información disponibles proporcionan el texto completo de artículos que se encuentran en la base de datos, mientras que otros sólo ofrecen el resumen o información de las citas. Por ejemplo, la Asociación Psicológica Americana ha creado una base de datos con textos completos llamada *PsycARTICLES* (http://www.apa.org/psycarticles/). Todos los artículos en esta base de datos provienen de revistas publicadas por la APA. Otras bases de datos con artículos de texto completo provienen de distintas fuentes. En ocasiones es atractivo limitarse a servicios que ofrecen textos completos, porque es muy sencillo obtener todo el artículo. Un problema de esta estrategia es que usted se limita únicamente a las revistas que están en las bases de datos de texto completo. Generalmente, ampliar su exploración es una buena idea, de modo que tenga mayores posibilidades de encontrar los artículos más relevantes para su tema. Incluso si el texto completo del artículo no está disponible por medio de la computadora, quizás pueda obtenerlo a través de otro recurso de su biblioteca.

Búsquedas en Internet

El recurso de información disponible más extendido es la riqueza de material gratuito en Internet. Servicios como Yahoo!, Google, About y AltaVista le permiten buscar a través de una variedad de materiales almacenados en Internet.

Ésta es una fuente maravillosa de información: cualquier búsqueda puede ayudarle a encontrar sitios dedicados al tema de su interés, artículos que la gente ha puesto a la disposición de los demás, revisiones de libros e incluso, discusiones en línea. Aun cuando es increíblemente fácil buscar (sólo escriba algo en el cuadro de diálogo y presione la tecla "Enter"), puede mejorar la calidad de sus búsquedas si aprende: 1) las diferencias en la forma en que cada servicio encuentra y almacena la información; 2) las reglas de la búsqueda avanzada, incluyendo cómo reducir las búsquedas y cómo encontrar frases exactas, y 3) formas para evaluar críticamente la calidad de la información que encuentre. También necesita asegurarse de registrar cuidadosamente el servicio de búsqueda, los términos que utilizó, la fecha y la ubicación exacta de cualquier sitio web que continuará utilizando para su investigación; esta información le será útil, siempre y cuando proporcione las referencias para los artículos que prepare.

Su propia biblioteca y una variedad de sitios y bibliotecas en Internet poseen información para evaluar la calidad de los datos encontrados en la red. Por ejemplo, usted querrá ver si el sitio web está asociado con una institución educativa, revisar los antecedentes de la persona u organización que financia el sitio, y verificar qué tan actual es la información.

Incluso con la expansión de las herramientas de búsquedas electrónicas fáciles de usar, probablemente también deseará realizar búsquedas por medio de algunos métodos "pasados de moda". Por ejemplo, cuando encuentre un artículo importante para su investigación, las referencias al final de éste pueden utilizarse como fuentes de ayuda para aprender más sobre el tema.

ANATOMÍA DE UN ARTÍCULO DE INVESTIGACIÓN

La búsqueda de literatura le ha ayudado a encontrar artículos de investigación que podrá leer. ¿Qué espera encontrar en esos documentos? Los artículos de investigación usualmente contienen *cinco* secciones: 1) un *resumen* o *abstract,* como los que se encuentran en los *Psycbological Abstracts;* 2) una *introducción* que explica el problema de investigación y las hipótesis específicas que se prueban; 3) una sección de *método* que describe con detalles los procedimientos exactos utilizados en el estudio; 4) una sección de *resultados,* en la que se presentan los hallazgos, y 5) una sección de *discusión* en la que el investigador puede especular sobre las aplicaciones más amplias de los resultados, en donde propone explicaciones alternativas para los resultados, discute razones por las que una hipótesis en particular no haya sido confirmada por los datos, y hace sugerencias para otras investigaciones sobre el mismo problema. Además de las *cinco secciones* principales, usted encontrará una lista de todas las referencias citadas.

Resumen

El resumen es una síntesis del reporte de investigación y usualmente está formado por no más de 120 palabras. Abarca información sobre las hipótesis, el

procedimiento y el patrón general de los resultados. Por lo general incluye poca información sobre la discusión del documento.

Introducción

En la introducción el investigador describe el problema de la investigación, explica con detalle los estudios realizados previamente y las teorías relevantes al problema; aclara sus predicciones específicas, comúnmente en forma de hipótesis. En otras palabras, el investigador introduce el estudio en un formato lógico que muestra cómo la teoría e investigaciones anteriores están conectadas con el problema de investigación presente y los resultados esperados.

Método

La sección del método está dividida en subsecciones, el autor determina cuántas subsecciones incluirá, dependiendo de la complejidad del diseño de investigación. En ocasiones, la primera subsección presenta un panorama general del diseño con el fin de preparar al lector para el material que se presenta a continuación. La siguiente subsección describe las características de los participantes. ¿Se trata de hombres o mujeres, o se utilizaron ambos sexos? ¿Cuál fue la edad promedio? ¿Cuántos participantes se incluyeron? Si el estudio incluyó seres humanos, es necesario explicar la forma en que fueron reclutados. La siguiente subsección describe con detalles los procedimientos empleados. Al describir los materiales presentados a los participantes, la forma en que se registraron sus conductas, etc., es importante que no se omita ningún detalle crucial. Estas particularidades hacen posible que el lector conozca la forma exacta en que se llevó a cabo el estudio, y permite que otros investigadores tengan la información necesaria para replicarlo. Es probable que se requieran otras subsecciones que describan el equipo o materiales de prueba que se utilizaron.

Resultados

En la sección de resultados el investigador presenta los hallazgos, habitualmen-:e de tres maneras. Primero, los describe en forma narrativa; por ejemplo, "la ubicación de los objetos tendía a olvidarse cuando era muy fácil de recordar o cuando se trataba de un lugar poco común para guardar ese objeto". En segundo lugar, los resultados se describen en lenguaje estadístico. En tercer lugar, el material se presenta en tablas y gráficas.

La terminología estadística de la sección de resultados puede parecer formidable. Sin embargo, la ausencia de conocimiento acerca de los cálculos no es un impedimento para entender el artículo o la lógica subyacente a la estadística. La estadística es sólo una herramienta que usa el investigador para evaluar los resultados del estudio.

Discusión

En la sección de la discusión, el investigador revisa el estudio desde varias perspectivas. ¿Los resultados confirman las hipótesis? Si es así, el autor debe dar todas las explicaciones posibles de los resultados y aclarar por qué una explicación es superior a otra. Si las hipótesis no fueron confirmadas, el autor debe sugerir algunas razones. ¿Qué problemas pueden haberse presentado en la metodología, en las hipótesis o en ambas? El investigador también puede comparar sus resultados con los de investigaciones anteriores sobre el mismo tema. En esta sección también se pueden incluir sugerencias sobre las posibles aplicaciones prácticas del estudio y sobre investigaciones futuras.

Usted debe leer tantos artículos de investigación como le sea posible para familiarizarse con la forma en que se presenta la información en los reportes. Conforme lea, desarrollará formas para procesar de manera eficiente la información de los artículos. Generalmente es mejor leer primero el resumen, después dar una hojeada al artículo para decidir si la información proporcionada le es útil. Si le es posible, lea el artículo cuidadosamente; observe las teorías y las hipótesis que se presentan en la introducción; anote cualquier cosa que le parezca poco clara o problemática en el método, y lea los resultados en función del material incluido en la introducción. Sea crítico al leer el documento, los estudiantes suelen dar las mejores críticas. Lo más importante es que, conforme lea más material sobre un tema, se familiarizará más con las variables que se estudian, los métodos empleados para analizarlas, los aspectos teóricos más importantes y los problemas que deben abordarse en futuras investigaciones. En resumen, estará generando sus propias ideas de investigación y planeando sus propios estudios.

Términos de estudio

Hipótesis
Predicción
Psychological Abstracts
PsycLIT, PsycINFO, PsycFIRST
Resumen
Revisión de la literatura
Science Citation Index (SCI)
Sección de discusión
Sección de introducción
Sección de resultados
Sección del método
Social Sciences Citation Index (SSCI)
Teoría

Preguntas de repaso

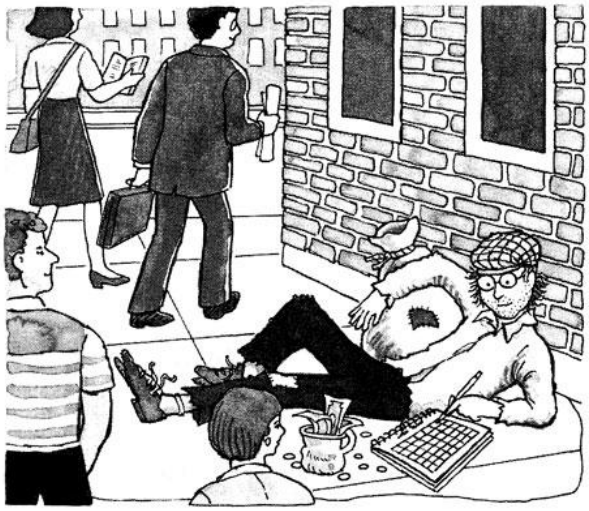
1. ¿Qué es una hipótesis? ¿Cuál es la diferencia entre una hipótesis y una predicción?
2. ¿Cuáles son las dos funciones de una teoría?
3. Describa las diferencias para encontrar investigaciones anteriores utilizando el *PsycINFO y* el método del "artículo clave" con el *Social Sciences Citation Index y* el *Science Citation Index.*

4. ¿Qué información comunica el investigador en cada una de las secciones de un artículo de investigación?

Actividades

1. Piense en al menos *cinco* proverbios "de sentido común" sobre el comportamiento (por ejemplo, "recurre al castigo y echa a perder al niño"; "de tal padre, tal hijo"; "la ausencia provoca más cariño"). Para cada una, elabore una hipótesis sugerida por el proverbio y una predicción que surja a partir de la hipótesis (basado en Gardner, 1988).
2. Elija una de las hipótesis formuladas en la actividad 1 *y* desarrolle una estrategia para encontrar investigaciones sobre el tema, utilice las bases de datos por computadora de su biblioteca.
3. Las teorías tienen dos propósitos: 1) organizar y explicar los acontecimientos observables, y 2) generar nuevos conocimientos al guiar la forma en que observamos estos sucesos. Identifique un patrón de conducta consistente en usted mismo o en alguien cercano a usted (por ejemplo, usted discute con su hermana los viernes en la noche). Genere dos posibles teorías (explicaciones) para este suceso (por ejemplo, debido a que usted trabaja muchas horas los viernes, generalmente se siente estresado y exhausto cuando llega a su casa. Debido a que su hermana tiene examen de química todos los viernes en la tarde y, puesto que no le va bien en el curso, se siente irritable los viernes). ¿De qué forma reuniría evidencias para determinar cuál de las explicaciones sería la correcta? ¿De qué manera cada explicación conduciría a distintos métodos para cambiar el patrón de conducta, ya sea disminuyendo o incrementando su ocurrencia?

3



Ética de la investigación

Experimento sobre obediencia de Milgram

El *Reporte Belmont*

Evaluación de riesgos y beneficios

Los riesgos en la investigación psicológica

Daño físico Estrés Pérdida de la privacidad y confidencialidad

Consentimiento informado

Formato de consentimiento informado

Aspectos de la autonomía

Aspectos sobre la información: ocultación de la información y engaño

¿Constituye el engaño un problema ético importante en la investigación psicológica?

La importancia del informe

Alternativas al engaño

Juego de roles

Estudios de simulación

Experimentos honestos

Justicia y selección de los participantes

Compromisos del investigador

Reglamentos federales y el Consejo institucional de revisión

Investigación exenta Investigación de riesgo mínimo Investigación de riesgo más que mínimo

Código ético de la APA

Investigación con participantes humanos

Utica e investigación con animales

Revisión de riesgos y beneficios

Falsificación

Términos de estudio

Preguntas de repaso

Actividades

Los aspectos *éticos* son cruciales cuando se planea, *realiza y* evalúa una investigación. En este capítulo exploraremos con detalle los aspectos éticos y examinaremos algunos lineamientos para abordar estos problemas.

EXPERIMENTO SOBRE OBEDIENCIA DE MILGRAM

Stanley Milgram realizó una serie de experimentos (1963,1964,1965) para estudiar el fenómeno de la obediencia a una figura de autoridad. Colocó un anuncio en el periódico local en New Haven, Connecticut, ofreciendo un pago de $4.50 a los hombres que participaran en un "estudio científico de memoria y aprendizaje", que se llevaría a cabo en la Universidad de Yale. Los participantes se re-portaron al laboratorio de Milgram, en Yale, donde conocieron a un científico vestido con una bata blanca, y a otro participante en el estudio, un hombre de mediana edad llamado "señor Wallace", quien era en realidad un cómplice del experimentador; pero los participantes no lo sabían. El científico explicó que el estudio examinaría los efectos del castigo sobre el aprendizaje. Una persona sería el "maestro" que administraría el castigo, y el otro sería el "aprendiz". El señor Wallace y el voluntario sacaron pedacitos de papel para determinar quién sería el maestro y quién el aprendiz. Sin embargo, la selección estaba manipulada, ya que el señor Wallace siempre era el aprendiz y el voluntario, el maestro.

El científico colocaba electrodos al señor Wallace y paraba al maestro frente a una impresionante máquina de choques eléctricos. El aparato tenía una serie de palancas que, según se le dijo al individuo, cuando fueran presionadas aplicarían descargas eléctricas al señor Wallace. La primera palanca tenía una etiqueta de 15 voltios; la segunda, una de 30 voltios; la tercera, una de 45 voltios, y así sucesivamente, hasta llegar a 450 voltios. Las palancas también tenían leyendas que indicaban "choque ligero", "choque moderado", hasta llegar a "peligro: choque grave", seguido por una X roja por arriba de los 400 voltios.

Se le indicó al señor Wallace que aprendiera una serie de palabras, ordenadas en pares; después se le aplicó una prueba para ver si podía identificar cuáles carían juntas. Cada vez que el señor Wallace cometía un error, el maestro debía aplicarle un choque eléctrico como castigo. El primer error debía castigarse con una descarga de 15 voltios, el segundo con uno de 30 voltios, etc.; cada vez que cometía un error, el aprendiz recibía un choque más intenso. En realidad, el señor Wallace nunca recibió una descarga eléctrica, pero los participantes del estudio no lo sabían. En el experimento, el señor Wallace cometía un error tras oro. Cuando el maestro "le aplicaba" un choque de alrededor de 120 voltios, el aprendiz empezaba a gritar de dolor y eventualmente exclamaba que ya no deseaba participar. ¿Qué sucedía si el maestro deseaba abandonar el estudio? Esto ocurrió, pues los participantes reales se alteraron notoriamente por el dolor que el señor Wallace parecía estar experimentando. El científico le decía al maestro que podía abandonar el estudio, pero lo motivaba a continuar por medio de comentarios verbales que enfatizaban la importancia de continuar con el experimento.

Supuestamente se trataba de un experimento sobre memoria y aprendizaje, aunque Milgram en verdad estaba interesado en saber si los participantes continuarían obedeciendo al experimentador, administrando choques eléctricos cada vez más fuertes al aprendiz. ¿Qué aconteció? Aproximadamente el 65 por ciento de los participantes continuó aplicando choques hasta los 450 voltios. El estudio de Milgram fue muy difundido, y los resultados desafían muchas de nuestras creencias sobre nuestra habilidad para resistirnos a la autoridad. El estudio de Milgram es importante y sus resultados tienen implicaciones para la comprensión de la obediencia en situaciones reales, como en el holocausto en la Alemania nazi y el suicidio masivo de Jonestown (véase Miller, 1986). ¿Qué se puede decir de los aspectos éticos en el estudio de Milgram? ¿Cómo podemos decidir si el estudio de Milgram o cualquier otro es ético?

EL *REPORTE BELMONT*

Los lineamientos éticos actuales para los investigadores médicos y del comportamiento tienen su origen en el *Reporte Belmont.* Principios y Lineamientos Éticos para la Protección de los Sujetos Humanos en la Investigación. Comisión Nacional para la protección de los sujetos humanos en investigación biomédica y conductual, 1979. Este reporte definió los principios y aplicaciones que han guiado los reglamentos más detallados y el Código de Ética de la Asociación Psicológica Americana. Los tres principios básicos son: el beneficio, el respeto por las personas (autonomía) y la justicia. Las formas de aplicar estos principios respectivamente, son la evaluación de los riesgos y los beneficios, el consentimiento informado y la selección de los sujetos. Éstas guiarán la siguiente explicación sobre los aspectos éticos de la investigación.

EVALUACIÓN DE RIESGOS Y BENEFICIOS

El principio del **beneficio** en el *Reporte Belmont* se refiere a la necesidad de que la investigación maximice los beneficios y minimice cualesquiera posibles efectos dañinos de la participación. En la mayoría de las decisiones que tomamos en nuestra vida, consideramos los riesgos (o costos) y beneficios relacionados con la decisión. En las decisiones que se refieren a la ética de la investigación, debemos calcular los riesgos y beneficios potenciales que pueden resultar; a esto se le llama *análisis de riesgo y beneficio.* Los principios éticos requieren que nos preguntemos si los procedimientos de una investigación minimizan los riesgos para los participantes.

Los riesgos potenciales para los participantes incluyen factores como el daño físico o psicológico y la pérdida de confidencialidad; estudiaremos estos factores con detalle. Además, puede considerarse el costo de *no* realizar el estudio si, de hecho, el procedimiento propuesto es la única forma de reunir datos potencialmente valiosos (véase Christensen, 1988). Los beneficios incluyen un

provecho directo para los participantes, como una ganancia educativa, la adquisición de nuevas habilidades o el tratamiento de un problema psicológico o médico. También puede haber beneficios materiales, como un pago en efectivo, cierto tipo de regalo e incluso la posibilidad de ganar un premio en una rifa. Otros beneficios menos tangibles incluyen la satisfacción obtenida al formar parte de una investigación científica y las aplicaciones benéficas potenciales de los hallazgos del estudio (por ejemplo, el conocimiento adquirido por medio de la investigación podría mejorar las prácticas educativas futuras, la psicoterapia o la política social). Como veremos más adelante, los reglamentos actuales con respecto de la realización de investigaciones con participantes humanos requieren de un análisis de riesgos y beneficios, antes de que éstas puedan ser aprobadas.

Los *riesgos en la investigación psicológica*

Regresemos a considerar la investigación de Milgram. El riesgo de experimentar estrés y daño psicológico es obvio. No es difícil imaginar los efectos de aplicar intensos choques eléctricos a un aprendiz reticente. Una película que tomó Milgram muestra a los participantes protestando, sudando e incluso riendo con nerviosismo mientras aplican los choques eléctricos. Usted se preguntará si está justificado someter a las personas a un experimento tan angustiante, *y* si la experiencia tuvo consecuencias a largo plazo para los voluntarios. Por ejemplo, ¿los participantes que obedecieron al experimentador sintieron remordimientos continuamente o empezaron a considerarse personas crueles e inhumanas? Después se incluye una defensa del estudio de Milgram; pero primero consideremos algunos procedimientos de investigación potencialmente inquietantes.

Daño físico Los procedimientos que pueden causar algún daño físico a los participantes son raros, pero posibles. Muchos tratamientos médicos caen en esta categoría, por ejemplo, la administración de una droga como el alcohol o la cafeína o privar a las personas de sueño durante un largo periodo. Los riesgos en procedimientos como éstos requieren de mucho cuidado para que sean aceptables éticamente. Además, deben existir beneficios claros de la investigación, que superen los riesgos potenciales.

Estrés El estrés psicológico es más común que el físico. Por ejemplo, a los participantes se les puede avisar que recibirán choques eléctricos sumamente intensos. En realidad, nunca reciben tales descargas; la variable de interés es el temor o ansiedad producidos durante el periodo de espera. La investigación realizada por Schachter (1959) en la que se utilizó un procedimiento como éste, demostró que la ansiedad produjo una necesidad de afiliación con los demás durante el periodo de espera.

En otro proceso que provoca estrés psicológico, a los participantes se les da retroalimentación desfavorable sobre su personalidad o habilidades. Los investigadores interesados en la autoestima suelen efectuar a los sujetos una prueba

de personalidad o de habilidades que en realidad es falsa; después de la aplicación de la prueba presentan una evaluación que disminuye o incrementa la autoestima, cuando le indican al participante que tiene un rasgo desfavorable de personalidad o una puntuación baja en las habilidades.

El hecho de preguntarles a las personas acerca de sucesos traumáticos o desagradables de sus vidas también puede causar estrés en algunos participantes. Así, investigaciones que piden a los sujetos que piensen en la muerte de uno de sus padres, del esposo o de un amigo, o en los recuerdos de un desastre vivido, pueden disparar una reacción de estrés.

Cuando es posible que se genere estrés, se debe preguntar si se han tomado todas las medidas de seguridad para ayudar a los participantes a afrontar el estrés. Generalmente existe una sesión de "información" después del estudio, diseñada en parte para tratar cualquier problema potencial que pueda haber surgido durante la investigación.

Pérdida de la privacidad y confidencialidad Otro riesgo es la pérdida de la intimidad y confidencialidad esperadas. Los investigadores deben tener cuidado en proteger la privacidad de los individuos. Como mínimo, los especialistas deben proteger la privacidad manteniendo todos los datos en un lugar seguro. Este tema se vuelve especialmente importante cuando se estudian temas como la conducta sexual, el divorcio, la violencia familiar o el abuso de drogas; en estos casos, los científicos quizás requieran plantear a las personas preguntas sensibles sobre su vida privada. Es extremadamente importante que las respuestas a estas preguntas sean confidenciales. En la mayoría de los casos, las respuestas son completamente anónimas, no existe forma de conectar la identidad de cualquier persona con los datos. Esto sucede, por ejemplo, cuando se administrar, cuestionarios a grupos de personas y no se pide información que podría utilizarse para identificar a un individuo (como el nombre, el registro del seguro social o el número telefónico). En otros casos, como en una entrevista personal en la que la identidad de la persona podría conocerse, el investigador debe planea: cuidadosamente las formas para codificar los datos, almacenarlos y explicar los procedimientos a los participantes, de modo que no exista ninguna duda con respecto al anonimato de las respuestas.

En algunas investigaciones existe una necesidad real de poder identificar a los participantes de manera individual. Esto sucede cuando se estudia a los sujetos en múltiples ocasiones en diferentes momentos, o cuando se da una retro-alimentación personal, como el puntaje de una prueba. En estos casos, los investigadores deben crear formas para identificar a los individuos, pero separan de la información sobre su identidad de los datos reales. Así, si alguien ve los cuestionarios o los archivos de datos computarizados, éstos no pueden relacionarse con individuos específicos.

En ciertos casos, los riesgos relacionados con la pérdida de la confidencialidad son tan grandes que los investigadores pueden solicitar un Certificado de confidencialidad del Departamento de salud y servicios humanos de

Estados Unidos (U.S. Department of Health and Human Services). Es apropiado obtener este certificado cuando los datos podrían ser objeto de un proceso legal.

Otro aspecto de la privacidad se refiere a la observación encubierta de la conducta. En algunos estudios, los investigadores realizan observaciones del comportamiento en lugares públicos. Observar a las personas en tiendas de autoservicio o en sus vehículos no parece implicar ningún problema ético. Sin embargo, ¿qué sucede si un investigador desea observar la conducta en ambientes más privados o en formas que podrían violar la privacidad de los individuos? (Véase Wilson y Donnerstein, 1976.) Por ejemplo, ¿sería ético esculcar en la basura de la gente u observarla en baños públicos? En un estudio, Middlemist, Knowles y Matter (1976) midieron los tiempos de la conducta de orinar en varones, su inicio y su duración, en los baños de una universidad. El propósito de la investigación era estudiar el efecto del espacio personal en una medida de activación fisiológica (los momentos de orinar). Se observó a los estudiantes mientras estaban solos o con un cómplice del experimentador, que permanecía en el apartado contiguo o en uno más alejado en el mismo baño. La presencia y la cercanía del cómplice tuvo el efecto de retrasar la conducta de orinar y de acortar su duración. De muchas formas, se trata de un análisis interesante; además, esta situación suele ser común para los hombres. No obstante, uno puede cuestionarse si se justifica la invasión de la privacidad (Koocher, 1977). Los investigadores, a su vez, pueden argumentar que por medio de estudios piloto y discusiones con los participantes potenciales determinaron que los problemas éticos del estudio eran mínimos (Middlemist *et. al.*, 1977). Middlemist y sus colaboradores emplearon un método para determinar si un procedimiento es ético, propuesto originalmente por Berscheid, Baron, Dermer y Libman (1973). El juego de roles se utiliza para reunir evidencia sobre la percepción que los participantes tienen de un experimento potencial. Si los participantes en el juego de roles indican que sí participarían en el experimento, al menos se ha manejado una objeción para el engaño.

CONSENTIMIENTO INFORMADO

El principio del *Reporte Belmont*, sobre el *respeto por las personas*, o **autonomía,** establece que los participantes deben ser tratados como seres autónomos; son capaces de tomar una decisión voluntaria con respecto a su participación o no en la investigación. En este sentido se aplica el **consentimiento informado:** se debe proporcionar a los probables participantes toda la información que pueda influir en su decisión. Se les debe informar sobre los propósitos de la investigación, los riesgos y beneficios de tomar parte en ella y su derecho a negarse o finalizar su participación en el estudio. Después de esto pueden aceptar o negarse libremente a participar en la investigación.

Formato de consentimiento informado

Generalmente se proporciona a los participantes algún tipo de formato de consentimiento informado, que contiene la información que todo sujeto necesita para tomar su decisión. Lo más común es que se trate de un formato impreso que el participante lee y firma. Es importante que los individuos comprendan la información del formato. Ha habido casos en los que el formato es demasiado técnico o que está cargado de términos legales, de modo que los participantes no comprenden en absoluto lo que están firmando. Los formatos del consentimiento deben redactarse en lenguaje sencillo y directo, que evite términos técnicos (por lo general, en un nivel entre primero y tercero de secundaria; en su mayoría, los procesadores ofrecen información sobre el nivel con el revisor de gramática). Para lograr que el formato sea más fácil de entender, debe evitarse redactarlo en primera persona, sino que se debe proporcionar la información como si el investigador estuviese teniendo una conversación con el participante. Así, el formato podría decir:

> *La participación en este estudio es voluntaria. Usted puede negarse a participar sin ser penalizado.*

en lugar de:

> *Comprendo que la participación en este estudio es voluntaria. Puedo negarme a participar sin ser penalizado.*

El primer párrafo proporciona información al participante de forma directa, en segunda persona (usted), mientras que el segundo tiene un tono legal que puede ser más difícil de entender. Finalmente, si los participantes no hablan español, debe existir una versión traducida del formato. La figura 3.1 proporciona una lista de verificación de un formato de consentimiento informado.

Aspectos *de la autonomía*

El consentimiento informado parece ser bastante sencillo, sin embargo, existen aspectos importantes que deben tomarse en cuenta. El primero se refiere a la falta de autonomía. ¿Qué sucede cuando los participantes no poseen la capacidad para tomar una decisión libre e informada para participar voluntariamente? Las poblaciones especiales, como los menores de edad, los pacientes de hospitales psiquiátricos o los adultos con deterioro cognoscitivo, requieren de precauciones también especiales. Por ejemplo, cuando se pide la participación de menores de edad, generalmente se necesita el formato de consentimiento escrito y firmado por uno de los padres o tutores, además de un acuerdo por parte del menor, el cual se denomina formalmente *asentimiento.* La División de Psicología del Desarrollo de la Asociación Psicológica Americana y la Sociedad

Asegúrese de que el formato de consentimiento informado incluya lo siguiente:

_____ La aseveración de que se les pide a los participantes que formen parte de un estudio de investigación.

_____ La explicación de los propósitos de la investigación en un lenguaje claro.

_____ La duración esperada de la participación de los sujetos.

_____ La descripción de los procedimientos.

_____ La descripción de los riesgos o incomodidades razonablemente previsibles y las medidas de seguridad para minimizar los riesgos.

---------- La descripción de los beneficios para el individuo o para otros que pueden esperarse, de manera razonable, de la investigación.

_____ Si es el caso, indicar los procedimientos o cursos alternativos apropiados de tratamiento, en caso de que existan, que puedan ser ventajosos para el individuo.

_____ La descripción del grado en que se mantendrá la confidencialidad de los registros que identifican al individuo, si éstos existieran.

_____ Si se ofrece un incentivo, su descripción y los requisitos para obtenerlo; además, la descripción de las implicaciones de una decisión de no continuar participando.

_____ Información sobre la forma de hacer contacto para plantear dudas acerca del estudio (generalmente contacto de tipo telefónico con el investigador, el consejero y la oficina del Consejo de revisión institucional).

_____ La afirmación de que la participación es voluntaria, que el negarse a participar no implica penalización alguna ni la pérdida de los beneficios a los que de otra forma el sujeto tiene derecho, y que el individuo puede suspender su participación en cualquier momento, sin sanción alguna ni la pérdida de los beneficios a los que de otra forma tiene derecho.

_____ El formato debe estar impreso en letra no menor de 11 puntos (y no en "impresión fina").

_____ El formato no debe incluir términos técnicos y debe estar escrito en un nivel de entre primero y tercero de secundaria.

_____ El formato no debe redactarse en primera persona (frases como "yo comprendo...", deben evitarse).

Las investigaciones con "alto riesgo" y las que incluyen procedimientos médicos pueden necesitar información adicional. Puede encontrar más información sobre la elaboración de un formato de consentimiento informado en sitios de Internet de universidades y del gobierno federal de Estados Unidos, como Guías para el consentimiento informado (Tips on Informed Consent) del Departamento de servicios humanos y de salud (Department of Health and Human Services): http://ohrp.osophs.dhhs.gov/humansubjects/guidance/ictips.htm.

FIGURA 3.1
Lista del formato de consentimiento informado.

para la Investigación del Desarrollo del Niño han establecido sus propios lineamientos para la investigación ética con niños.

Otra amenaza para la autonomía es la coerción. Cualquier procedimiento que limite la libertad de un individuo para aceptar, es potencialmente coercitivo. Por ejemplo, un supervisor que pide a sus empleados que llenen una encuesta durante una junta de personal o un profesor que pide a los estudiantes en una clase que participen en un estudio para poder aprobar el curso, ambos están ejerciendo una gran presión sobre los participantes potenciales. Los empleados pueden creer que el supervisor los sancionará de alguna forma si no colaboran, también se arriesgan al ridículo si se niegan a hacerlo frente a sus compañeros de trabajo. En ocasiones, los beneficios son tan grandes que se vuelven coercitivos. Por ejemplo, un prisionero puede creer que su participación le dará mayores privilegios o incluso la libertad condicional. Los investigadores deben considerar estos aspectos y asegurarse de preservar la autonomía.

Aspectos sobre la información: ocultación de la información y engaño

Tal vez se le habrá ocurrido a usted que sería imprudente proporcionar toda la información sobre el estudio. Demasiada información podría invalidar los resultados del estudio, por ejemplo, los investigadores acostumbran retener información sobre las hipótesis o la condición particular en la que participa un individuo (véase Sieber, 1992). Generalmente es aceptable retener información cuando ésta no afecta la decisión de participar, y cuando los participantes recibirán dicha información más tarde, durante una sesión de informe después de que el estudio se haya completado. La mayoría de las personas que participan voluntariamente en investigaciones psicológicas no esperan que se les revelen todos los aspectos del estudio antes de su participación. Sin embargo, sí esperan toda la información después de finalizar el estudio. El informe será descrito después de considerar el engaño, que es un aspecto más problemático.

Quizás también se le habrá ocurrido que existen procedimientos de investigación en los que el consentimiento informado no es necesario, o que incluso no es posible. Si elige observar el número de grupos de estudio con individuos del mismo sexo y de distintos sexos en su biblioteca, probablemente no necesite anunciar su presencia y obtener el permiso de cada uno de ellos. Si estudia el contenido de las autodescripciones de las personas que escriben a un servicio de citas en Internet, ¿necesita ponerse en contacto con cada una de ellas para incluir la información en su estudio? Cuando se planea una investigación, es importante asegurarse de que existen buenas razones para prescindir del consentimiento informado.

El engaño ocurre cuando existe una deformación activa de la información. El experimento de Milgram ilustra dos tipos de engaño. Primero, hubo un engaño con relación a los propósitos del estudio. Los participantes aceptaron formar parte de un estudio sobre memoria y aprendizaje, pero en realidad participaron en un estudio acerca de la obediencia. ¿Quién podría imaginar que un experimento sobre memoria y aprendizaje (después de todo, el título suena inofen-

sivo) implicaría la aplicación de choques eléctricos dolorosos de gran intensidad a otra persona? Los participantes del experimento de Milgram no sabían a qué se estaban prestando.

El estudio de Milgram se realizó antes de que el consentimiento informado se convirtiera en una rutina; no obstante, se puede imaginar que el formato de consentimiento de Milgram haría, de manera inapropiada, que los participantes aceptaran formar parte de un estudio sobre la memoria. También se les había dicho que eran libres de abandonar el estudio en cualquier momento. ¿Es posible que el procedimiento de consentimiento informado afectara los resultados del estudio? El saber que la investigación está diseñada para estudiar la obediencia, probablemente altere el comportamiento de los participantes. A pocos de nosotros nos gusta considerarnos obedientes, y probablemente trataríamos de probar que no lo somos. Las investigaciones indican que el consentimiento informado puede, de hecho, sesgar las respuestas de los participantes, por lo menos en ciertas áreas. Por ejemplo, las investigaciones sobre elementos estresantes como el ruido o el hacinamiento, han demostrado que la sensación de tener control" sobre uno de estos factores reduce su impacto negativo. Si usted sabe que puede detener un ruido fuerte *y* molesto, el ruido produce menor estrés que cuando se percibe como incontrolable. Estudios realizados por Gardner (1978) *v* Dill, Gilden, Hill *y* Hanslka (1982), han demostrado que los procedimientos del consentimiento informado incrementan la percepción de control en experimentos de estrés y, por lo tanto, pueden afectar a las conclusiones obtenidas a partir de ellos.

También es posible que el procedimiento del consentimiento informado produzca un sesgo en la muestra. En el experimento de Milgram, si los participantes hubiesen sabido que se les pediría aplicar choques severos a la otra persona, algunos se habrían negado a participar en él. Por lo tanto, podríamos limitar nuestra capacidad para generalizar los resultados únicamente a aquellos 'tipos" que aceptaron participar. Si esto fuese así, ¡cualquiera podría decir que la conducta de obediencia observada en el experimento de Milgram se dio sencillamente debido a que las personas que aceptaron formar parte eran, en primer lugar, unos sádicos!

El estudio de Milgram también ejemplifica un tipo de engaño en el que los participantes pasan a formar parte de una serie de acontecimientos establecidos para los propósitos del estudio. Un cómplice del experimentador desempeñó el papel de otro participante del estudio; Milgram creó una realidad para el participante en la que podría observarse la obediencia ante una autoridad. Este tipo de engaño ha sido el más común en las investigaciones de la psicología social; es menos frecuente en áreas de la psicología experimental, tales como la percepción, el aprendizaje, la memoria y el desempeño motor en seres humanos. Incluso en estas áreas, los investigadores pueden utilizar una historia encubierta para que el experimento parezca plausible y atractivo (por ejemplo, decirle a los participantes que están leyendo historias periodísticas actuales para un estudio sobre la amenidad de la lectura, cuando el verdadero propósito es examinar errores de memoria o esquemas de organización).

El problema del engaño no está limitado a la investigación de laboratorio. Los procedimientos en los que los observadores encubren sus propósitos, presencia o identidad, también son engañosos. Por ejemplo, Humphreys (1970) estudió el comportamiento de homosexuales varones que frecuentaban baños públicos (llamados salones de té). Humphreys no participó en actividades homosexuales, sino que fungió como vigilante para avisar a los demás sobre posibles intrusos. Además de observar las actividades en el salón de té, Humphreys anotó el número de las placas de los visitantes al salón. Posteriormente, obtuvo las direcciones de los hombres, se disfrazó y visitó sus hogares para entrevistarlos. Ciertamente, su procedimiento es una forma de estudiar la homosexualidad, pero con un gran componente de engaño.

A lo largo de los años, psicólogos como Kelman (1967), Ring (1967), Rubin (1970, 1985) y Ortmann y Hertwig (1997) han criticado el empleo del engaño. Kelman, Ortmann y Hertwig consideran que cualquier tipo de engaño es moral-mente inaceptable; sencillamente es incorrecto confundir a las personas en cualquier forma. Otra objeción al uso del engaño es que daña la reputación del campo. Rubin y Ring también argumentan que los investigadores en ocasiones diseñan engaños elaborados, debido a la notoriedad que provocan; Rubin los describe como "bromistas salvajes del laboratorio". Este autor ha señalado un procedimiento en el que un estudiante de posgrado, que fungió como experimentador, hizo que los participantes creyeran que habían roto una valiosa pieza de equipo. En otro ejemplo, Ortmann y Hertwig escriben un estudio en el que un cómplice se enfrentó con un participante y lo insultó con un nombre despectivo, con la intención de provocarle enojo.

¿Constituye el engaño un problema ético importante en la investigación psicológica?

Muchos psicólogos consideran exagerado el problema del engaño (Bröder, 1998; Kimmel, 1998; Korn, 1998; Smith y Richardson, 1985). Bröder argumenta que los ejemplos extremos de engaño elaborado, citados por estos críticos, son poco comunes. Además, existe evidencia de que a los estudiantes universitarios que participan en investigaciones no les preocupa el engaño y que, de hecho, pueden disfrutar los experimentos que incluyen engaños (Christensen, 1988).

Algunos investigadores han intentado evaluar el uso del engaño durante las últimas décadas, desde los experimentos de Milgram en los años sesenta, para ver si el engaño elaborado en realidad se ha vuelto menos común. Puesto que la mayor parte del empleo de este tipo de engaño surge en las investigaciones de psicología social, los intentos por enfatizar este aspecto se han enfocado en esta área de la psicología. Gross y Fleming (1982) revisaron 691 estudios de psicología social, publicados en los años sesenta y setenta. Aun cuando la mayoría de los estudios en los años setenta continuaban utilizando el engaño, éste implicaba principalmente historias falsas.

¿Ha continuado la tendencia a abandonar el engaño? Sieber, Iannuzzo y Rodríguez (1995) examinaron los estudios publicados en *el Journal ofPersonality*

and Social Psycbology en 1969, 1978, 1986 y 1992. La cantidad de estudios que utilizaron alguna forma de engaño disminuyó de 66 por ciento en 1969 a 47 por ciento en 1978, y a 32 por ciento en 1986; pero se incrementó nuevamente a 47 por ciento en 1992. La gran caída de 1986 puede deberse al aumento, en ese año, en el número de estudios sobre temas como la personalidad, que no requieren el uso del engaño. Además, en 1992 el consentimiento informado se describía de manera más explícita que en años anteriores, y era más probable que el informe se mencionara en los años posteriores a 1969. Sin embargo, las historias falsas de encubrimiento continúan utilizándose con frecuencia. Korn (1997) también ha concluido que el empleo del engaño está disminuyendo en la psicología social.

Existen tres razones principales para el decremento en el tipo de engaños elaborados observados en el estudio de Milgram. En primer lugar, más investigadores se han interesado en las variables cognoscitivas que en las emociones, por lo cual utilizan métodos similares a los que se emplean en las investigaciones de la memoria y la psicología cognoscitiva. En segundo lugar, el nivel general de preocupación por los aspectos éticos, como se describe en este capítulo, ha hecho que los investigadores realicen sus estudios de otras maneras (más adelante se describen algunas alternativas al engaño). En tercer lugar, los comités de ética en las universidades y colegios ahora revisan las propuestas de investigaciones con mayor cuidado, de manera que el engaño elaborado únicamente se aprueba cuando el estudio es importante y no hay procedimientos alternativos disponibles (los consejos de revisión ética se describen posteriormente en este capítulo).

LA IMPORTANCIA DEL INFORME

El **informe** se da después de que se termina el estudio. Es una oportunidad para que el investigador aborde aspectos como la ocultación de información, el engaño y los efectos potencialmente dañinos de la participación.

Si los participantes fueron engañados de alguna forma, el investigador debe explicar por qué fue necesario hacerlo. Si el estudio, de alguna forma, alteró el estado físico o psicológico de algún participante, como sucede en un experimento que produce estrés, el investigador debe asegurarse de que el participante ya está "calmado" y que se siente cómodo por haber participado. Si existe la necesidad de que el participante reciba información adicional o de que hable con alguien más acerca del estudio, el investigador debe proporcionar el acceso a estos recursos. Es importante que los participantes terminen el estudio sin resentimientos hacia el campo de la psicología y que, incluso, puedan obtener cierta comprensión de su propio comportamiento o personalidad.

El informe también ofrece una oportunidad para que los investigadores expliquen los propósitos del estudio y comuniquen a los participantes el tipo de resultados que se espera; también se pueden comentar las implicaciones prácticas de éstos e incluso discutirse. En algunos casos, los investigadores pueden

ponerse en contacto con los participantes posteriormente para informarles acerca de los resultados reales del estudio. Así, el propósito del informe es tanto educativo como ético.

¿Es suficiente el informe para eliminar cualesquiera efectos negativos, cuando se incluye el estrés y el engaño elaborado? Volvamos de nuevo a la investigación de Milgram. Este autor llegó muy lejos para proporcionar una profunda sesión de informe. A los participantes que fueron obedientes, se les dijo que su comportamiento era normal y que no habían actuado de forma distinta a los otros participantes. Se les hizo conscientes de la fuerte presión situacional a la que fueron expuestos, y se hicieron esfuerzos para reducir cualquier tensión que hubieran experimentado. Se les aseguró que en realidad no habían aplicado ningún choque eléctrico, y hubo una reconciliación amistosa con el cómplice, el señor Wallace. Milgram también envió por correo un reporte con los hallazgos de su investigación a los participantes, y les preguntó acerca de sus reacciones al experimento. Las respuestas mostraron que el 84 por ciento de ellos estaban satisfechos de haber participado y el 74 por ciento dijo que se había beneficiado con la experiencia. Únicamente el 1 por ciento afirmó sentirse arrepentido de haber participado. Cuando un psiquiatra entrevistó a los participantes un año después, no se pudieron detectar efectos dañinos por la participación. Solamente podemos concluir que el informe tuvo los efectos que se buscaban. Otros investigadores que han realizado más estudios sobre la ética del experimento de Milgram, llegaron a la misma conclusión (Ring, Wallston y Corey, 1970).

Otra investigación sobre el informe también concluyó que éste es una forma eficiente para manejar el engaño y otros aspectos éticos que surgen en las investigaciones (Smith, 1983; Smith y Richardson, 1983). Sin embargo, como se señaló antes, el engaño continúa siendo un tema polémico. Entonces, quizás usted se pregunte si existen alternativas al engaño.

ALTERNATIVAS AL ENGAÑO

Después de criticar el uso del engaño en la investigación, Kelman (1967) exhortó a la creación de procedimientos alternativos. Uno que él sugiere es el juego de roles, y otras opciones incluidas son los estudios de simulación (una variante del juego de roles) y los experimentos "honestos".

Juego de roles

En un procedimiento de **juego de roles,** el experimentador describe una situación a los participantes y después les pregunta cómo responderían a ésta. En ocasiones, se les pide a los participantes que digan cómo se comportarían en la situación; otras veces, se les pide que predigan el comportamiento de los participantes reales en una situación como ésa. No está claro si estas dos instrucciones producen alguna diferencia en los resultados.

El juego de roles no suele ser considerado como una alternativa satisfactoria para el engaño (Freedman, 1969; Miller, 1972). Un problema es que la simple lectura de la descripción de una situación no involucra con gran profundidad a los participantes, no forman parte de una situación real. Además, debido a que el experimentador da a los participantes una descripción completa de la situación, su hipótesis puede volverse obvia para ellos. Cuando las personas descubren la hipótesis, pueden tratar de comportarse de una forma consistente con ésta. Los detalles del experimento que pueden dar información sobre las hipótesis a los participantes se denominan "características de demanda". El problema de las características de demanda se describe con detalle en el capítulo 9.

El defecto más grave del juego de roles es que, no importa qué resultados se obtengan, los críticos siempre podrán afirmar que éstos habrían sido diferentes si los participantes hubieran estado en una situación real. Esta crítica se basa en el supuesto de que la gente no siempre es capaz de predecir con precisión su propia conducta o la de otros. Esto es verdad, en particular, cuando se involucran conductas indeseables como el conformismo, la obediencia o la agresión. Por ejemplo, si Milgram hubiese utilizado el procedimiento del juego de roles, ¿cuántas personas cree usted que habrían pronosticado que serían completamente obedientes? De hecho, Milgram le pidió a un grupo de psiquiatras que predi-leran los resultados de su estudio y encontró que, incluso, estos expertos no podían anticipar con exactitud lo que sucedería. Surgiría un problema similar si se le pidiera a la gente que vaticinara si ayudaría a alguien necesitado, la mayoría de nosotros probablemente sobreestimaríamos nuestras tendencias altruistas.

Estudios de simulación

Un tipo diferente de juego de roles implica la simulación de una situación real. Un estudio de simulacro que impresionó a Kelman (1967) es la Simulación entre naciones, en la que los participantes jugaron a ser los líderes de naciones y los investigadores observaron los procesos de negociación, resolución de problemas, entre otros. Simulaciones como ésta puede producir un importante invo-lucramiento personal, tal como lo haría cualquiera que se haya visto absorbido profundamente por un juego de Internet con otros usuarios o incluso por el juego de Monopolio.

También pueden existir problemas éticos con las simulaciones. Un ejemplo dramático es el Experimento de la prisión de Stanford, realizado por Zimbardo (1973; Haney y Zimbardo, 1998). Este autor estableció una prisión fingida en el sótano del edificio de psicología en la Universidad de Stanford. Después reclutó estudiantes universitarios a quienes se les pagaron $15 diarios por representar el papel de prisioneros o guardias durante un periodo de dos semanas. A los guardias se les vistió con uniformes, lentes oscuros y toletes. A los prisioneros se les asignaron números y utilizaron gorros hechos con medias de nylon para simular cortes de pelo de prisioneros y para reducir los sentimientos de in-dividualidad. Los participantes se involucraron profundamente en sus papeles,

tanto que Zimbardo tuvo que detener la simulación después de seis días debido al comportamiento cruel de los "guardias" y a las reacciones de estrés de los "prisioneros". Se trató únicamente de una simulación, ya que los participantes sabían que no eran realmente prisioneros ni guardias. Sin embargo, se involucraron tanto en sus papeles, que el experimento produjo niveles de estrés más altos que casi cualquier otro experimento que uno pudiera imaginar. Afortunadamente, el experimento de Zimbardo es un caso poco común, la mayoría de los estudios de simulación no implican los aspectos éticos observados en este estudio en particular.

Experimentos honestos

Rubin (1973) describió lo que él llamó estrategias experimentales "honestas", ninguna de ellas incluye el juego de roles. La primera estrategia es una en la que los participantes están totalmente conscientes de los propósitos de la investigación. En un estudio realizado por Byrne, Ervin y Lamberth (1970), los investigadores dijeron a estudiantes universitarios que estaban interesados en la eficacia de las citas por computadora. Los especialistas utilizaron un programa de cómputo para formar parejas de estudiantes, hombres y mujeres, que tenían actitudes similares o distintas. Cada pareja tuvo una cita breve en el *campus* universitario, y después los investigadores midieron la atracción que sintieron el uno por el otro. Las parejas similares se sentían más atraídas entre sí que las parejas distintas. El estudio no incluyó engaños ni distorsiones de los propósitos de la investigación.

Una segunda estrategia honesta se utiliza en situaciones en las que los programas buscan explícitamente cambiar el comportamiento de la gente. Ejemplos citados por Rubin incluyen los programas educativos, invitaciones a la salud, campañas de caridad y políticas, y solicitudes de voluntarios. En estas situaciones las personas son conscientes de que alguien intenta cambiar su conducta. Por ejemplo, las personas se pueden exponer de manera voluntaria a una invitación para dejar de fumar. Los investigadores pueden, entonces, estudiar los efectos de dicha invitación mientras manipulan variables como la cantidad de temor provocado (Leventahl, 1970).

Rubin también aseveró que muchos experimentos de campo involucran procedimientos honestos. A diferencia de los experimentos de laboratorio como el estudio de Milgram, en el que el comportamiento se estudia en un ambiente de laboratorio, los experimentos de campo introducen la manipulación experimental en un contexto natural. Rubin cita un estudio en el que un investigador observaba a conductores de automóviles mientras esperaban el cambio de la luz roja (Ellsworth, Carlsmith y Henson, 1972). Los automovilistas cruzaron la intersección con mayor velocidad que conductores en una situación control, en la que no había un experimentador observando. Este experimento en particular no parece implicar problemas éticos, todos hemos tenido la experiencia de ser observados. Los investigadores únicamente aplicaron métodos experimentales para estudiar de forma sistemática esta situación. Gran parte de la investigación

en escenarios de campo es efectivamente honesta. Los investigadores observan la conducta de las personas en lugares públicos y en situaciones cotidianas. No obstante, el hecho de que un experimento se realice en el campo no significa que no implique aspectos éticos. Por ejemplo, ¿cuáles son las implicaciones éticas (y legales) de quitarle el tiempo a vendedores de automóviles, cuando se finge ser un cliente, para estudiar diferentes tipos de ofertas de precios o de exponer a los pasajeros del Metro a una persona que se desmaya entre cada estación? (Véase Silverman, 1975.)

La última estrategia honesta comentada por Rubin incluye situaciones en las que un acontecimiento que ocurre de forma natural representa una oportunidad de investigación: "la naturaleza, el destino, el gobierno y otras fuerzas inalterables con frecuencia imponen su voluntad sobre las personas, de forma aleatoria o no sistemática". Por ejemplo, ciertos investigadores pudieron estudiar los efectos del hacinamiento, cuando la escasez de vivienda para estudiantes obligó a que la Universidad de Rutgers asignara aleatoriamente a los nuevos estudiantes a dormitorios hacinados y a otros no hacinados (Aiello, Baum y Gormley, 1981). Baum, Gachtel y Schaeffer (1983) estudiaron los efectos estresantes asociados con desastres nucleares, compararon personas que vivían cerca de la planta nuclear de Three Mile Island, con otras que vivían cerca de una planta nuclear sin daños, o de una planta convencional que funcionaba con carbón. La ciencia depende de la posibilidad de replicar los resultados, de modo que es notable que se haya obtenido el mismo patrón de resultados después del ataque terrorista del 11 de septiembre (Schlenger *et al,* 2002). Más de 2 000 residentes adultos de las ciudades de Nueva York, Washington D.C., y otras áreas metropolitanas a lo largo de Estados Unidos, respondieron un cuestionario para identificar el trastorno por estrés postraumático (TPEPT), con el fin de determinar la incidencia de este padecimiento. El 11.2 por ciento de los residentes de Nueva York presentó el TPEPT, en contraste con el 2.7 por ciento de los residentes de Washington y el 3.6 por ciento de los individuos que vivían en otras áreas urbanas. Experimentos naturales como éstos constituyen valiosas fuentes de datos.

JUSTICIA Y SELECCIÓN DE LOS PARTICIPANTES

El tercer principio ético definido en el *Reporte Belmont* es el de la **justicia.** El principio de justicia enfatiza la igualdad cuando se reciben los beneficios de la investigación, así como la responsabilidad de aceptar los riesgos. La historia de la investigación médica incluye demasiados ejemplos de estudios de alto riesgo, realizados con individuos seleccionados por su falta de poder y marginación dentro de la sociedad. Uno de los más terribles es el Estudio de sífilis de Tuskegee, en el que 399 afroestadounidenses pobres de Alabama no recibieron tratamiento contra la sífilis, para así poder rastrear los efectos a largo plazo de esta enfermedad (Reverby, 2000). Este estudio se llevó a cabo de 1932 a 1972, cuando sus detalles se hicieron públicos. La indignación provocada por el hecho de que este

estudio se hubiese realizado y de que los sujetos fueran afroestadounidenses que no sospechaban nada, motivó a los científicos a crear reglas éticas en la investigación médica y del comportamiento.

El principio de justicia requiere que los investigadores tomen en cuenta los aspectos de la igualdad. Cualquier decisión de incluir o excluir a ciertas personas de un estudio de investigación debe justificarse con bases científicas. Por lo tanto, si se utiliza la edad, el grupo étnico, el género u otro criterio para seleccionar a los participantes, debe existir un fundamento científico.

COMPROMISOS DEL INVESTIGADOR

Los investigadores establecen varios "contratos" implícitos con los participantes durante el transcurso de un estudio. Por ejemplo, si los participantes están de acuerdo con presentarse para un estudio a una hora específica, el investigador debe estar ahí. Los científicos nunca mencionan el tema de la puntualidad, aunque los participantes lo señalan cuando se les pregunta acerca de las obligaciones del investigador (Epstein, Suedfeld y Silverstein, 1973). Si los investigadores prometen enviar un resumen de los resultados a los participantes, deben hacerlo. Si se estableció que los participantes recibirán un crédito por su participación, el investigador debe avisar de inmediato al instructor de que la persona formó parte del estudio. Éstos son "pequeños detalles", pero muy importantes, para mantener la confianza entre los participantes y los investigadores.

REGLAMENTOS FEDERALES Y EL CONSEJO INSTITUCIONAL DE REVISIÓN

El *Reporte Belmont* delinió los temas de la ética en la investigación. El Departamento de salud y servicios humanos de Estados Unidos (U.S. Department of Health and Human Services, HHS) emitió las reglas y los reglamentos para la protección de los participantes en investigaciones humanas. Bajo estas reglas (Department of Health and Human Services, 2001), todas las instituciones que reciben fondos federales deben tener un **Consejo institucional de revisión (CIR),** que sea responsable de la revisión de las investigaciones realizadas dentro de la institución. El CIR es una instancia de supervisión local compuesta por un mínimo de cinco individuos, por lo menos uno de sus miembros debe ser externo a la institución. Todas las universidades en Estados Unidos que reciben fondos federales tienen un CIR, además, la mayoría de los departamentos de psicología poseen su propio comité de revisión de investigaciones (Chastain y Landrum, 1999). Todos los estudios llevados a cabo por profesores, estudiantes y personal asociado a la institución, son revisados de alguna forma por el CIR. Esto incluye las investigaciones que se hacen en otros lugares, como una escuela, una agencia comunitaria, un hospital o por medio de Internet.

Los reglamentos federales para la supervisión de investigaciones de los CIR continúan evolucionando. Por ejemplo, en la actualidad todos los investigadores deben cumplir ciertos requisitos educativos. La mayoría de las universidades exigen a los estudiantes y profesores que respondan uno o más tutoriales en línea, sobre la ética de investigación, para cubrir estos requisitos. Usted puede encontrar fácilmente los tutoriales en una búsqueda de Internet.

Los reglamentos del HHS también categorizaron las investigaciones de acuerdo con la cantidad de riesgo que implican. Este concepto de riesgo se incorporó posteriormente en el Código de Ética de la Asociación Psicológica Americana.

Investigación exenta

La investigación *sin riesgos* está exenta de revisiones. Los estudios que incluyen únicamente cuestionarios anónimos, encuestas y pruebas educativas, están exentos, así como la observación naturalista en lugares públicos, cuando no existe una amenaza para el anonimato. La investigación documental en la que se estudian datos del dominio público o en la que no puede identificarse a los participantes, también está exenta. Este tipo de investigación tampoco requiere de un consentimiento informado. Sin embargo, debe haber un mecanismo institucional que determine qué investigaciones quedan exentas en realidad. Los investigadores no pueden decidir por sí mismos cuáles estudios están exentos, los CIR de las instituciones formulan el procedimiento que permite a un investigador solicitar el estatus de exenta para una investigación.

Investigación de riesgo mínimo

Un segundo tipo de actividad de investigación se denomina *de riesgo mínimo*, lo cual significa que las posibilidades de dañar a los participantes no superan a los riesgos que existen en la vida diaria o en exámenes físicos y psicológicos de rutina. Cuando se realiza una investigación de riesgo mínimo, las medidas de seguridad elaboradas son menos importantes y la aprobación del CIR es sólo un trámite. Algunas de las actividades de investigación consideradas de riesgo mínimo son: 1) el registro rutinario de datos fisiológicos de participantes adultos (por ejemplo, tomar el peso, las pruebas de agudeza sensorial, los electrocardiogramas, los electroencefalogramas, la ecografía diagnóstica y los registros de voz); observe que no se incluyen registros que podrían implicar la invasión de la privacidad; 2) el ejercicio moderado de voluntarios sanos, y 3) la investigación del comportamiento individual o grupal o de las características de individuos, como los estudios de percepción, cognición, teoría del juego o elaboración de pruebas en las que el investigador no manipula el comportamiento de los participantes, así como las investigaciones que no producen estrés en los participantes.

Investigación de riesgo más que mínimo

Cualquier procedimiento de investigación que ponga a los participantes en un riesgo mayor que el mínimo, está sujeto a una revisión profunda por el CIR. Se requiere de un consentimiento informado completo y de otras medidas de seguridad para garantizar su aprobación.

Los investigadores que planean realizar un experimento deben enviar una solicitud al CIR. La petición requiere la descripción de los riesgos y los beneficios, de los procedimientos para minimizar los riesgos, la redacción exacta del formato del consentimiento informado, la manera en que se informará a los participantes y los procedimientos para mantener la confidencialidad. Incluso, después que un proyecto ha sido aprobado, existe una revisión continua. Si se trata de un proyecto de largo plazo, éste será revisado al menos una vez al año. Si se hace cualquier cambio en el procedimiento, los investigadores deben obtener la aprobación del CIR.

Con las reglas del HHS y la revisión de investigaciones que realiza el CIR, los derechos y la seguridad de los participantes humanos están protegidos adecuadamente. En este punto debemos señalar que los investigadores y los miembros del consejo de revisión suelen ser muy cuidadosos con respecto de lo que se considera ético. De hecho, diversos estudios han demostrado que los alumnos que han participado en pruebas de investigación son más benévolos en sus juicios con respecto a la ética de los experimentos que los investigadores o los miembros de los CIR (Epstein *et al.*, 1973; Smith, 1983; Sullivan y Deiker, 1973). Además, individuos que han formado parte de investigaciones que utilizaron el engaño reportan que no les importa y han evaluado la experiencia de manera positiva (Christensen, 1988).

CÓDIGO DE ÉTICA DE LA APA

Los psicólogos reconocen los aspectos éticos que hemos comentado, y la Asociación Psicológica Americana (APA) es líder en la formulación de principios y estándares éticos. Los *Principios éticos de los psicólogos y el Código de conducta*, conocido como *Código de ética* se revisó en 2002 y está disponible en http://www.apa.org/ethics (Asociación Psicológica Americana 2002a). El preámbulo de este *Código de ética* establece que:

> Los psicólogos están comprometidos a incrementar el conocimiento científico y profesional sobre el comportamiento y la comprensión que las personas tienen de sí mismas y de los demás, y a utilizar este conocimiento para mejorar la condición de los individuos, las organizaciones y la sociedad. Los psicólogos respetan y protegen los derechos civiles y humanos y la importancia central de la libertad de búsqueda y expresión en la investigación, la enseñanza y la publicación. Luchan por ayudar al público a desarrollar juicios y elecciones informadas con respecto a la conducta humana. Para lograrlo, desempeñan diversos papeles, como investigadores,

> educadores, diagnosticadores, terapeutas, supervisores, consultores, administradores, interventores sociales y testigos expertos. Este *Código de ética* proporciona un conjunto común de principios y estándares con base en los cuales los psicólogos construyen su trabajo profesional y científico.

Cinco principios generales se relacionan con el beneficio, la responsabilidad, la integridad, la justicia y el respeto por los derechos y dignidad de los demás. Diez fundamentos éticos enfatizan aspectos específicos relativos a la conducta de los psicólogos en la enseñanza, la investigación, la terapia, la consejería, las pruebas otros papeles y responsabilidades profesionales. Nos enfocaremos en el Están-dar ético 8: investigación y publicación.

INVESTIGACIÓN CON PARTICIPANTES HUMANOS

Las secciones del Estándar ético 8 que abordan directamente la investigación con participantes humanos son:

> ***8.01 Aprobación institucional***
>
> Cuando se requiere de la aprobación institucional, los psicólogos proporcionan información precisa acerca de los propósitos de su investigación y obtienen una aprobación antes del realizar el estudio. Conducen el estudio de acuerdo con el protocolo de investigación aprobado.

> ***8.02 Consentimiento informado para investigar***
>
> *a)* Para obtener el consentimiento informado, tal como se requiere en el estándar 3.10, Consentimiento informado, los psicólogos informan a los participantes acerca de: 1) el propósito de la investigación, su duración esperada y los procedimientos; 2) su derecho a negarse a participar y a abandonar la investigación una vez que su participación haya iniciado; 3) las consecuencias previsibles de negarse o abandonar; 4) los factores previsibles que se pueden esperar razonablemente que influyan en su disposición para participar, como los riesgos potenciales, la incomodidad o los efectos adversos; 5) cualesquiera posibles beneficios de la investigación; 6) los límites de la confidencialidad; 7) los incentivos para participar, y 8) a quién contactar en caso de tener preguntas sobre la investigación y acerca de los derechos de los participantes. Ellos dan oportunidad a los futuros participantes de plantear preguntas y de recibir respuestas (véanse también los Estándares 8.03, Consentimiento informado para la Grabación de voces e imágenes en investigaciones; 8.05, Exención del

consentimiento informado para investigación; y 8.07, Engaño en la investigación). *b)* Los psicólogos que realizan investigación con intervención, utilizando tratamientos experimentales, aclaran a los participantes al *inicio* de la investigación: 1) la naturaleza experimental del tratamiento; 2) los servicios que estarán o no disponibles para el (los) grupo(s) control, *si* es apropiado; 3) los medios con los que se hará la asignación al tratamiento y al (a los) grupo(s) control; 4) las alternativas de tratamiento disponibles si un individuo no desea participar en la investigación o desea abandonarla una vez que haya iniciado, y 5) la compensación o los costos en dinero por la participación y, de ser el caso, *si* habrá un reembolso de parte del participante o de un tercero, (véase también el Estándar 8.02a, Consentimiento informado para investigar).

8.03 Consentimiento informado para grabar voces e imágenes en investigaciones

Los psicólogos obtienen el consentimiento informado de los participantes de la investigación antes de grabar sus voces o imágenes para la recolección de datos, a menos que: 1) el estudio consista únicamente de observaciones naturales en lugares públicos, y que no se anticipe que la grabación será utilizada de manera que pueda causar daño personal o permita la identificación del individuo; o 2) que el diseño de la investigación incluya el engaño y se obtenga el consentimiento informado durante el informe (véase también el Estándar 8.07, Engaño en la investigación).

8.04 Cliente/paciente, estudiante y participantes de investigación subordinados

a) Cuando los psicólogos realizan investigaciones con clientes o pacientes, estudiantes o subordinados como participantes, toman las medidas para proteger a los participantes potenciales de las consecuencias adversas de negarse a participar o abandonar el estudio.

b) Cuando la participación en una investigación es requisito para algún curso o representa la oportunidad para obtener más créditos, se debe dar al participante potencial la opción de actividades alternativas equivalentes.

8.05 Exención del consentimiento informado para investigar

Los psicólogos pueden exentar el consentimiento informado únicamente: 1) cuando se supone, de manera razonable, que la investigación no provocará malestar o daño e incluye: (a) el estudio de prácticas educativas

normales, curriculares o métodos para el manejo en el salón de clases, realizados en ambientes educativos; (b) uso únicamente de cuestionarios anónimos, observaciones naturales o investigación documental, en las que la revelación de las respuestas no pone a los participantes en riesgo de responsabilidad criminal o civil o de daño a su situación financiera, su empleo o su reputación, y se protege la confidencialidad; o (c) el estudio de factores relacionados con la eficacia de empleos u organizaciones, llevados a cabo en ambientes organizacionales, en los que el empleo del participante no está en riesgo y se protege la confidencialidad; o 2) cuando las leyes o las reglas federales o institucionales lo permiten.

8.06 Ofrecimiento de incentivos para la participación en investigaciones

a) Los psicólogos hacen esfuerzos razonables para evitar ofrecer incentivos *económicos o* de otro tipo, excesivos o inapropiados, por la participación en investigaciones, cuando esa clase de incentivos puede coaccionar la participación.

b) Cuando ofrecen servicios profesionales como incentivos para la participación en investigaciones, los psicólogos aclaran la naturaleza de los servicios, así como los riesgos, obligaciones y limitaciones (véase también el Estándar 6.05 Intercambio con los clientes/pacientes).

8.07 Engaño en ¡a investigación

a) Los psicólogos no realizan un estudio que implique el engaño, a menos que hayan determinado que el uso de técnicas con engaño esté justificado por su importante valor potencial *científico,* educativo o aplicado, y siempre que no se disponga de procedimientos alternativos *efectivos* que no recurran al engaño.

b) Los psicólogos no engañan a los participantes potenciales con respecto a investigaciones en las que se puede esperar, dentro de lo razonable, que causen dolor físico o malestar emocional grave.

c) Los psicólogos explican a los participantes, tan pronto como sea posible, cualquier engaño que constituya una característica integral del diseño y de la conducción de un experimento, de preferencia al final de su participación, pero antes de concluir la recolección de datos, y permiten a los participantes retirar sus datos (véase también el Estándar 8.08, Informe).

8.08 Informe

a) Los psicólogos ofrecen una pronta oportunidad a los participantes para obtener la información apropiada acerca de la naturaleza, los

resultados y las conclusiones de la investigación, y toman medidas razonables para corregir cualquier distorsión que éstos tengan con respecto a lo que conocen los psicólogos.

b) Si los valores científicos o humanos justifican el retraso o la ocultación de esta información, los psicólogos toman medidas razonables para reducir el riesgo de daño.

c) Cuando los psicólogos se dan cuenta de que los procedimientos de la investigación han dañado a un participante, toman medidas razonables para minimizar el daño.

Estos estándares complementan las reglas del HSS y del *Reporte Belmont.* Asimismo, enfatizan la importancia del consentimiento informado como parte fundamental de una práctica ética. Sin embargo, no siempre es posible obtener un consentimiento informado completo, y en ocasiones el engaño es necesario. En estos casos se incrementan las responsabilidades del investigador con los participantes. Obviamente, las decisiones con respecto de lo que se considera ético o poco ético no son sencillas; no existen reglas absolutas.

ÉTICA E INVESTIGACIÓN CON ANIMALES

Aun cuando este capítulo se ha centrado en la ética de la investigación con seres humanos, seguramente usted sabe que los psicólogos en ocasiones realizan investigaciones con animales. Los animales se utilizan por diversas razones, el investigador puede controlar cuidadosamente las condiciones ambientales de los animales, estudiarlos durante un largo periodo y observar su conducta 24 horas al día, en caso necesario. Los animales también se utilizan para probar los efectos de drogas y para estudiar los mecanismos fisiológicos y genéticos que sub-yacen a la conducta. Aproximadamente 7 por ciento de los artículos en los *Psychological Abstracts* de 1979 estudiaron animales (Gallup y Suárez, 1985), y los datos indican que la cantidad de investigación realizada con animales ha ido disminuyendo de manera estable (Thomas y Blackman, 1992). Lo más común es que los psicólogos trabajen con ratas y ratones, y en menor proporción con aves. Según una encuesta sobre la investigación con animales en psicología, más del 95 por ciento de los animales empleados en la investigación fueron ratas, ratones y aves (véase Gallup y Suárez, 1985).

En años recientes, grupos que se oponen a la investigación con animales en medicina, psicología, biología y otras ciencias, se han vuelto más vehementes y activos. Organismos que defienden los derechos de los animales han manifestado sus protestas en las convenciones de la Asociación Psicológica Americana. y miembros de estos grupos han robado animales de laboratorios de investigación en diversas ciudades. Estos grupos también han presionado para que se promulguen leyes que prohíban toda la investigación con animales.

Los científicos argumentan que la investigación con animales beneficia a los seres humanos y señalan una cantidad considerable de descubrimientos que

no hubieran sido posibles sin este tipo de investigación (Miller, 1985; Carroll y Overmier, 2001). Además, las organizaciones que defienden los derechos de los animales suelen exagerar la cantidad de investigación que involucra dolor o sufrimiento (Coile y Miller, 1984).

Plous (1996a, 1996b) realizó una encuesta nacional con psicólogos y estudiantes de psicología, sobre las actitudes hacia el uso de animales en investigación y educación (disponible en http://www.apa.org/science/anguide.html). La actitud de los psicólogos y de los estudiantes fue bastante similar. En general existe un apoyo a la investigación con animales: el 72 por ciento de los estudiantes la apoya, el 18 por ciento se opone y el 10 por ciento duda (no obstante, los psicólogos apoyan con mayor "fuerza" la investigación animal que los estudiantes). Además, el 68 por ciento considera que la investigación con animales es necesaria para el progreso de la psicología. Sin embargo, existe cierta ambivalencia e incertidumbre con respecto del uso de animales: cuando se les preguntó si los animales en las investigaciones psicológicas reciben un trato humano, el 12 por ciento de los estudiantes dijo "no" y el 44 por ciento se manifestó "con duda". Además, la investigación que incluye ratas o pichones se considera más positiva que la que emplea perros o primates, a menos que los estudios sean estrictamente de observación. Finalmente, las mujeres tienen opiniones menos positivas hacia la investigación con animales que los hombres. Plous concluyó que la investigación con animales en psicología mantendrá su importancia para el campo, pero muy probablemente seguirá disminuyendo su proporción con respecto al total de las investigaciones que se realizan.

La investigación con animales es, en realidad, muy importante y continuará siendo necesaria para estudiar muchos tipos de preguntas de investigación. Es crucial reconocer que existen leyes y lineamientos éticos estrictos que gobiernan la investigación con animales y los procedimientos de enseñanza en los que se emplean estos individuos. Dichos reglamentos hablan de la necesidad de lugares, alimentación, limpieza y cuidados de salud apropiados. Asimismo, especifican que la investigación debe evitar cualquier tipo de crueldad en la forma de dolor innecesario en los animales. Además, las instituciones en las que se realiza investigación con animales deben tener un *Comité institucional para el cuidado y uso de animales* (CICUA), compuesto de al menos un científico, un veterinario y un miembro de la comunidad. El CICUA tiene la obligación de revisar los procedimientos de investigación con animales y de asegurar que se cumplan todos los reglamentos (véase Holden, 1987). Esta sección del Código de ética es especialmente importante aquí:

8.09 ***Cuidado y uso humano de los animales en la investigación***

a) Los psicólogos adquieren, cuidan, emplean y disponen de los animales de acuerdo con las leyes y reglamentos federales, estatales y locales vigentes, y bajo estándares profesionales.

b) Psicólogos entrenados en métodos de investigación y con experiencia en el cuidado de animales de laboratorio, supervisan todos los

procedimientos que incluyen animales y son responsables de asegurar un trato adecuado para su comodidad, salud y trato humano.

c) Los psicólogos se aseguran que todos los individuos que estén bajo su supervisión y que utilicen animales hayan recibido instrucción en métodos de investigación y en el cuidado, mantenimiento y manejo de las especies empleadas, hasta el punto en que su papel lo requiera (véase también el Estándar 2.05, Delegación del trabajo a otros).

d) Los psicólogos realizan esfuerzos razonables para minimizar la incomodidad, las infecciones, las enfermedades y el dolor de los animales.

e) Los psicólogos utilizan un procedimiento que somete a los animales al dolor, estrés o privaciones únicamente cuando no se dispone de un procedimiento alternativo, y cuando la meta está justificada por su potencial valor científico, educativo o aplicado.

f) Los psicólogos realizan procedimientos quirúrgicos bajo anestesia apropiada, y utilizan técnicas para evitar infecciones y minimizar el dolor durante y después de la cirugía.

g) Cuando es apropiado terminar con la vida de un animal, los psicólogos proceden con rapidez y esforzándose por minimizar el dolor, de acuerdo con los procedimientos aceptados.

La APA también ha elaborado detallados *Lineamientos para la conducta ética en el cuidado y uso de animales* (Asociación Psicológica Americana, 2002b). Es claro que los psicólogos están preocupados por el bienestar de los animales que se emplean en la investigación. Sin embargo, éste probablemente continuará siendo un tema polémico.

REVISIÓN DE RIESGOS Y BENEFICIOS

Ahora usted está familiarizado con los temas éticos que enfrentan los investigadores que estudian el comportamiento humano y animal. Cuando se toman decisiones sobre la ética de la investigación, es necesario tomar en cuenta los múltiples factores asociados con el riesgo para los participantes. ¿Existe riesgo de daño psicológico o de la pérdida de la confidencialidad? ¿Quiénes son los participantes de la investigación? ¿Qué tipos de engaño, si los hay, se utilizan en el procedimiento? ¿Cómo se obtendrá el consentimiento informado? ¿Qué procedimientos de informe se utilizan? También es necesario ponderar los beneficios directos de la investigación para los participantes, así como la importancia científica del estudio y los beneficios educativos que recibirán los estudiantes que realizan la investigación para una clase o como un requisito para graduarse Estas decisiones no son fáciles. Considere el estudio descrito anteriormente en el que un cómplice varón insulta a un participante varón. Este estudio, conducido por Cohén, Nisbett, Bowdle y Schwarz (1996), comparó las reacciones de

estudiantes universitarios que vivían al norte de Estados Unidos, con la de estudiantes que habitaban en el sur de este país. El propósito era investigar si los hombres del sur habían desarrollado una "cultura de honor" que los hacía responder de forma agresiva al ser insultados. De hecho, los estudiantes del norte respondieron poco al insulto, mientras que los del sur respondieron con elevados indicadores fisiológicos y cognoscitivos de ira. El hecho de que gran parte de la violencia en el mundo sea manifestada por hombres que responden al percibir un insulto a su honor, hace que este tema sea especialmente importante para la sociedad. ¿Considera usted que los beneficios potenciales del estudio para la sociedad y la ciencia superan a los riesgos involucrados en el procedimiento?

Obviamente, el CIR que revisó este estudio concluyó que los investigadores habían minimizado los riesgos para los participantes lo suficiente como para que los beneficios superaran a los costos.

Si al final usted decide que los costos sobrepasan a los beneficios, debe concluir que el estudio no puede ser realizado de esa forma. Quizás existan procedimientos alternativos que podrían emplearse para hacerlo aceptable. Si los beneficios superan los costos, probablemente usted decidirá que la investigación debe llevarse a cabo. Sus cálculos pueden diferir de los de otra persona, razón por la cual es buena idea tener consejos éticos de revisión. Una revisión adecuada de las propuestas de investigación disminuye las posibilidades de que se aprueben estudios que carecen de ética.

FALSIFICACIÓN

Debemos señalar otros dos elementos del Código de ética:

8.10 Reporte de los resultados de investigación

a) Los psicólogos no fabrican datos (véase también el Estándar 5.01a, Evitación de aseveraciones falsas o engañosas).

b) Si los psicólogos descubren errores significativos en los datos publicados, toman medidas razonables para corregirlos en una corrección, retracción, fe de erratas u otros medios de publicación apropiados.

8.11 Plagio

Los psicólogos no presentan porciones del trabajo o datos de otros autores como propios, incluso si citan ocasionalmente la fuente del trabajo o datos de otros autores.

La fabricación de datos constituye un fraude. Debemos ser capaces de creer en los resultados de investigación reportados, de otra forma, la base completa del

método científico como medio de conocimiento se ve amenazada. De hecho, a pesar de que el fraude puede darse en muchos campos, probablemente sea más grave en dos áreas: la ciencia y el periodismo. Esto se debe a que la ciencia y el periodismo son campos en los que se asume que los reportes escritos son descripciones precisas de hechos reales. No existen agencias contables independientes que verifiquen las actividades de los científicos y los periodistas.

Los casos de fraude en el campo de la psicología son considerados muy graves (véase Hostetler, 1987; Riordan y Marlin, 1987); aunque por fortuna son muy raros (Murray, 2002). Quizás el caso más famoso es el de Sir Cyril Burt, quien reportó que las puntuaciones del CI de gemelos idénticos criados de forma separada eran sumamente similares. Los datos se utilizaron para apoyar el argumento de que las influencias genéticas sobre el CI son extremadamente importantes. Sin embargo, Kamin (1974) señaló algunas irregularidades en los datos de Burt. Una cantidad de correlaciones entre los diferentes conjuntos de gemelos eran exactamente iguales hasta las milésimas, lo cual es virtualmente imposible en términos matemáticos. Esta observación condujo al descubrimiento de que algunos supuestos colaboradores de Burt en realidad no habían trabajado con él o simplemente los había inventado. No obstante, irónicamente los "datos" de Burt se acercaban a lo que han reportado otros investigadores que han estudiado las puntuaciones del CI de gemelos.

En la mayoría de los casos, el fraude se detecta cuando otros científicos no pueden replicar los resultados de un estudio. Cuando empezaron a surgir muchas preguntas sobre la investigación publicada por Karen Ruggiero, una psicóloga de la Universidad de Texas, debido a dificultades para replicarla, la doctora Ruggiero renunció a su cargo y se retractó de los hallazgos de su investigación (Murray, 2002). En ocasiones un colega que ha trabajado con el investigador detecta el fraude. Por ejemplo, Stephen Breuning fue culpable de falsear datos que demostraban que los estimulantes podían emplearse para reducir la conducta hiperactiva y agresiva en niños con retardo severo (Byrne, 1988). En este caso, otro investigador que había trabajado de manera cercana con Breuning empezó a sospechar de los datos; entonces informó a la agencia federal que había financiado la investigación.

La razón más común para sospechar de fraude es cuando un hallazgo importante, o poco común, no puede replicarse. El fraude no es un problema principal en la ciencia, en parte debido a que los investigadores saben que otros leerán sus reportes y realizarán más investigaciones, incluyendo las replicaciones. Saben que su reputación y su carrera se verían gravemente dañadas si otros científicos concluyen que sus resultados son fraudulentos. Además, la posibilidad de detectar un fraude se ha incrementado en años recientes, ya que la disponibilidad de los datos ahora es más abierta: las reglas de la mayoría de las agencias financieras piden a los investigadores que otros científicos tengan acceso a sus datos.

Entonces, ¿por qué en ocasiones los investigadores cometen fraudes? Por un lado, algunas veces los científicos tienen un trabajo que los presiona de forma

extrema a producir resultados impresionantes. Claro que ésta no es una razón suficiente, ya que muchos investigadores mantienen estándares éticos altos bajo la misma presión. Otra razón es que los investigadores que sienten la necesidad de producir datos fraudulentos tienen un miedo exagerado al fracaso y una enorme necesidad de tener éxito y la admiración que lo acompaña. Si desea explorar más la dinámica del fraude, podría empezar con el libro de Hearnshaw (1979) sobre Sir Cyril Burt. El caso continúa rodeado de controversia: un volumen editado se llama *Cyril Burt: ¿Fraud or Framed?* (Macintosh, 1995). Sin embargo, la mayoría de los análisis concluyen que la investigación fue fraudulenta (Tucker, 1997).

Debemos hacer un señalamiento final: las acusaciones de fraude no deben hacerse a la ligera. Si usted no está de acuerdo con los resultados de algún autor en filosofía, política, religión u otras áreas, eso no quiere decir que sean fraudulentos. Incluso si no puede replicar los resultados, la razón puede recaer en aspectos de la metodología del estudio, más que en un fraude intencional. No obstante, el hecho de que el fraude pueda ser una posible explicación de los resultados, enfatiza la importancia de mantener registros cuidadosos y documentación de los procedimientos y resultados.

El plagio se refiere a hacer pasar el trabajo de otros como propio. Usted debe incluir citas apropiadas de sus fuentes. Su trabajo se verá fortalecido si cita las referencias (Harris, 2002).

En conclusión, debemos señalar que los lineamientos y reglamentos éticos están en constante evolución. El Código de ética de la APA y los reglamentos federales, estatales y locales se revisan de forma periódica. Los investigadores deben estar siempre al corriente de las políticas y procedimientos vigentes. En los siguientes capítulos revisaremos muchos procedimientos específicos para el estudio del comportamiento. Conforme lea acerca de estos procedimientos y los aplique a una investigación que le interese, recuerde que los aspectos éticos siempre son primordiales.

Términos de estudio

Análisis de riesgos y beneficios
Autonomía *(Reporte Belmont)*
Beneficio *{Reporte Belmont)*
CICUA
CIR
Código de ética
Confidencialidad
Consentimiento informado
Engaño
Estudios de simulación
Experimentos honestos
Fraude
Informe
Investigación de riesgo mínimo
Investigación exenta
Juego de roles
Justicia *(Reporte Belmont)*
Plagio
Reporte Belmont
Responsabilidad
Riesgo

Preguntas de repaso

1. Comente los principales temas éticos en la investigación del comportamiento, incluya los riesgos, los beneficios, el engaño, el informe, el consentimiento informado y la justicia. ¿Cómo pueden los investigadores ponderar la necesidad de realizar investigación contra la necesidad de procedimientos éticos?
2. ¿Por qué el consentimiento informado es un principio ético? ¿Cuáles son los problemas potenciales del consentimiento informado completo?
3. ¿Cuáles alternativas al engaño se describen en el texto?
4. Resuma los principios concernientes a la investigación con participantes humanos en el Código de ética de la APA.
5. ¿Cuál es la diferencia entre las actividades de investigación "sin riesgo" y "de riesgo mínimo"?
6. ¿Qué es un Consejo institucional de revisión?
7. Resuma los procedimientos éticos para la investigación con animales.
8. ¿Qué constituye un fraude? ¿Cuáles son algunas de las razones de su ocurrencia y por qué no sucede con mayor frecuencia?

Actividades

1. Considere el siguiente experimento, similar al realizado por Smith, Lingle y Brock (1978). Cada participante interactuó durante una hora con otra persona que en realidad era un cómplice. Después de esta interacción, ambas personas estuvieron de acuerdo con regresar una semana después para otra sesión. Cuando los participantes reales regresaron, se les informó que la persona que habían conocido la semana anterior había muerto. Entonces los investigadores midieron las reacciones de los individuos ante la muerte.
   a) Comente los aspectos éticos implicados en el experimento.
   b) ¿Violaría el experimentador los lincamientos estipulados en el Estándar 8 del Código de ética de la APA que se refiere a la investigación con participantes humanos? ¿En qué forma?
   c) ¿Qué métodos alternativos sugeriría usted para estudiar este problema (reacciones ante la muerte)?
   d) ¿Serían diferentes sus reacciones a este estudio si los participantes hubieran jugado con un niño, y después se les dijera que el niño había muerto?
2. En un procedimiento descrito en este capítulo se da a los participantes una retroalimentación falsa sobre un rasgo desfavorable de su personalidad o sobre un bajo nivel de habilidad. ¿Qué aspectos éticos están implicados en este procedimiento? Compare sus reacciones a este

inmediatamente que la investigación que se opone a su punto de vista debe ser incorrecta y quizás fraudulenta? ¿Debe poner atención a las políticas de la agencia que financió el estudio? ¿Qué debe hacer si no encuentra algo incorrecto en un estudio que no apoya su punto de vista? (Nota: Los ejemplos se basan en hallazgos de investigaciones reales. No se incluyen las referencias, si desea obtener detalles sobre las investigaciones descritas aquí, puede hacer una búsqueda utilizando los recursos electrónicos de su biblioteca.)

4

Estudio del comportamiento

Variables

Definiciones operacionales de variables

Relaciones entre variables
Relación lineal positiva Relación lineal negativa Relación curvilínea Ausencia de relación Relaciones y reducción de la incertidumbre

Métodos no experimentales *versus* experimentales
Método no experimental
Dirección de causa y efecto
El problema de la tercera variable
Método experimental
Control experimental
Aleatorización

Variables independiente y dependiente

Causalidad

Elección de un método: ventajas de diversos métodos
Artificialidad de los experimentos Consideraciones éticas y prácticas Variables de los participantes Descripción del comportamiento Predicciones exitosas del comportamiento futuro Ventajas de diversos métodos

Evaluación de la investigación: tres tipos de validez
Validez de constructo
Validez interna Validez externa

Términos de estudio

Preguntas de repaso

Actividades

En este capítulo exploraremos algunos de los aspectos y conceptos básicos que son necesarios para comprender el estudio científico del comportamiento. Empezaremos con la naturaleza de las variables, su medición y los tipos de relaciones que existen entre ellas. Después examinaremos métodos generales para estudiar estas relaciones.

VARIABLES

Una **variable** es cualquier acontecimiento, situación, conducta o característica individual que cambia. Ejemplos de variables que un psicólogo podría estudiar incluyen el desempeño en tareas cognoscitivas, la longitud de palabras, la densidad espacial, la inteligencia, el género, el tiempo de reacción, la tasa de olvido, la agresión, la credibilidad del orador, el cambio de actitudes, el enojo, el estrés, la edad, y la autoestima. Cada una de estas variables representa una clase general en la que varían aspectos específicos, que se denominan *niveles* o *valores* de la variable. Una variable debe tener dos o más niveles o valores. En algunas variables, los valores tienen propiedades numéricas o cuantitativas verdaderas. Suponga que el desempeño en una tarea corresponde a un puntaje en una prueba cognoscitiva de 50 preguntas, en la que los valores pueden variar desde cero respuestas correctas hasta 50 respuestas correctas; estos valores tienen propiedades numéricas. Los valores de otras variables no son de tipo numérico, sino que simplemente identifican categorías diferentes. Un ejemplo es el género, los valores para el género son hombres y mujeres. Éstos son distintos, pero no difieren en cantidad.

Las variables pueden clasificarse en cuatro categorías generales. Las *variable: situacionales* describen características de una situación o del ambiente: la longitud de las palabras que usted lee en un libro, la densidad espacial de un salón de clases, la credibilidad de una persona que trata de persuadirlo y el número de espectadores en una emergencia. Las *variables de respuesta* son las respuestas o conductas de individuos: como el tiempo de reacción, el desempeño en una taren cognoscitiva y la ayuda que se ofrece a una víctima en una emergencia. Las v*ariables de los participantes* o *sujetos* son las diferencias individuales; son las características de los individuos, entre ellas el género, la inteligencia y los rasgos de personalidad, como la extroversión. Finalmente, las *variables mediadoras* son procesos psicológicos que median los efectos de una variable situacional sobre una respuesta en particular. Como ejemplo, Darley y Latané (1968) encontraron que, cuando hay más espectadores ante una emergencia, existen menores posibilidades de que alguien ofrezca ayuda. Una variable mediadora, llamada difusión de la responsabilidad, se utilizó para explicar este fenómeno (véase la fi-gura 4.1). Cuando hay varios espectadores, la responsabilidad personal para ayudar se diluye entre todos, por lo que ninguno de ellos siente mucha responsabilidad. Sin embargo, cuando una persona es el único testigo de una emergencia, toda la responsabilidad recae sobre ella, lo cual aumenta las posibilidades de que proporcione ayuda. Por consiguiente, según Darley y Latané, el número de

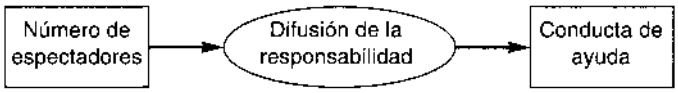


FIGURA 4.1
La difusión de la responsabilidad es una variable mediadora.

espectadores afecta la responsabilidad personal, que a su vez afecta la conducta de ayuda.

DEFINICIONES OPERACIONALES DE VARIABLES

En la investigación real, el investigador tiene que elegir un método con el cual estudiar las variables de interés. Es importante saber que una variable es un concepto abstracto que debe convertirse en formas concretas de observación o manipulación. Así, variables como "agresión", "desempeño en una tarea cognoscitiva", "cantidad de recompensa", "autoestima" e incluso "longitud de palabras", deben definirse en términos del método específico utilizado para medirlas o manipularlas. Los científicos se refieren a la **definición operacional** de una variable, es decir, una definición de la variable en términos de las operaciones o técnicas que el investigador utiliza para manipularla o medirla.

Las variables deben definirse operacionalmente para poder estudiarse de manera empírica. La variable "credibilidad del orador" podría conceptualizarse en dos niveles y definirse operacionalmente como un orador presentado a la audiencia como "merecedor del Premio Nobel" o como un "maestro sustituto de preparatoria". La variable "desempeño en una tarea cognoscitiva" podría definirse como el número de errores detectados en una prueba de lectura, durante un periodo de diez minutos.

También puede haber varios niveles de abstracción cuando se estudia una variable. Una variable como "longitud de palabras" es concreta y fácil de definir operacionalmente en términos del número de letras o sílabas, pero aun así, deben seleccionarse las palabras exactas para el estudio. El concepto de "estrés" es muy general y más abstracto. Cuando los investigadores estudian el estrés, pueden enfocarse en cualquier cantidad de factores estresantes: hacinamiento, ruido, problemas de salud graves, agotamiento por el trabajo, etc. Un investigador interesado en el estrés probablemente elegiría un agente estresante para su estudio y haría definiciones operacionales de éste. Después, realizaría investigaciones concernientes tanto al factor estresante específico como al *concepto* más general de estrés. El punto es que los investigadores siempre deben transformar las variables en operaciones específicas para manipularlas o medirlas.

La tarea de definir operacionalmente una variable obliga a los científicos a hablar de conceptos abstractos en términos concretos. El proceso puede dar como resultado que el investigador se dé cuenta de que la variable es demasiado vaga como para ser estudiada. Esto no necesariamente indica que el concepto carezca de significado, sino que una investigación sistemática no es posible hasta que el concepto se defina operacionalmente. Una vez que se encuentra una

definición operacional, el progreso en la comprensión de un fenómeno psicológico suele depender del desarrollo de tecnología cada vez más sofisticada. Por ejemplo, el concepto de "actividad cerebral" no es nuevo. El estudio de la forma en que la actividad cerebral se relaciona con el comportamiento se *vio* facilitado, primero, por el desarrollo de las técnicas de registro electrofisiológico y, más recientemente, por la tecnología de las imágenes cerebrales.

Las definiciones operacionales también nos ayudan a comunicar nuestras ideas a los demás. Si alguien desea decirme algo acerca de la agresión, necesito saber exactamente el significado de este término, ya que existen muchas formas de definirlo operacionalmente. Por ejemplo, la agresión podría definirse como: 1) la cantidad y duración de choques eléctricos aplicados a otra persona; 2) el número de veces que un niño golpea un juguete inflable; 3) el número de ocasiones en que un niño pelea con otros durante el recreo; 4) las estadísticas de homicidios obtenidas de los registros policiacos; 5) un puntaje en un cuestionario de agresividad; e incluso, 6) el número de veces que el bateador es golpeado en los lanzamientos durante los juegos de béisbol. La comunicación con otra persona es más fácil si coincidimos exactamente en lo que queremos decir cuando utilizamos el término *agresión* en el contexto de nuestra investigación.

En raras ocasiones existe un método único e infalible para definir operacionalmente una variable. Se puede disponer de una variedad de métodos, cada uno con ventajas y desventajas. Los investigadores deben decidir cuál utilizar, dependiendo del problema de estudio en particular, de las metas de la investigación y de otros aspectos como la ética y los costos. Para ilustrar qué tan complejo puede ser el desarrollo de una definición operacional de una variable, considere las opciones a las que se enfrenta un investigador que estudia el hacinamiento. Podría analizar los efectos que tiene el hacinamiento en estudiantes universitarios, en un experimento de laboratorio cuidadosamente controlado. Sin embargo, el foco de interés del investigador podría ser las secuelas a largo plazo del hacinamiento; para ello, quizás fuese buena idea observar las consecuencias que tiene el hacinamiento en animales de laboratorio, como ratas. El investigador podría examinar las repercusiones que tiene a largo plazo el hacinamiento sobre la agresión, la alimentación, la conducta sexual y la conducta maternal. Pero, ¿qué sucede si el investigador desea estudiar variables cognoscitivas o sociales, tales como el desempeño intelectual o la interacción familiar? En este caso, el científico podría decidir estudiar a personas que viven hacinadas y compararlas con otras que viven en circunstancias menos hacinadas. Puesto que ningún método es perfecto, la comprensión completa de cualquier variable implica estudiarla por medio de varias definiciones operacionales. A lo largo del libro se revisarán diversos métodos.

RELACIONES ENTRE VARIABLES

Muchas investigaciones estudian la relación entre dos variables, que es la forma general en que distintos valores de una variable se asocian con diferentes valores

de otra. Es decir, ¿varían sistemáticamente los niveles de las dos variables en conjunto? Conforme la edad aumenta, ¿se incrementa también la cantidad de juego cooperativo? ¿Ver violencia en la televisión resulta en mayor agresividad? ¿Está relacionada la credibilidad del orador con el cambio de actitudes?

Recuerde que algunas variables poseen valores numéricos verdaderos, mientras que los niveles de otras variables sencillamente son categorías diferentes. Estas diferencias se explicarán con mayor detalle en el capítulo 5. Con el propósito de describir relaciones entre variables, iniciaremos explicando las relaciones en que ambas variables tienen propiedades numéricas verdaderas.

Cuando ambas variables poseen valores a lo largo de una escala numérica, muchas "formas" diferentes pueden describir su relación. Primero nos enfocaremos en las cuatro relaciones más comunes que existen en la investigación: la **relación lineal positiva,** la **relación lineal negativa,** la **relación curvilínea** y, por supuesto, la situación en que *no hay relación* entre las variables. Estas relaciones se representan mejor por medio de gráficas lineales que muestran la forma en que los cambios en una variable se acompañan de los cambios de una segunda variable. Las cuatro gráficas de la figura 4.2 muestran los cuatro tipos de relaciones.

Relación lineal positiva

En una relación lineal positiva, el incremento en los valores de una variable se acompaña del aumento en los valores de la segunda variable. Anteriormente describimos una relación positiva entre la credibilidad del comunicador y la persuasión, mayores niveles de credibilidad están asociados con un mayor cambio de las actitudes. Considere otra variable del comunicador, la velocidad del discurso. ¿Son más persuasivos los "oradores rápidos"? En un estudio realizado por Smith y Shaffer (1991), diversos estudiantes escucharon un discurso expresado a velocidad lenta (144 palabras por minuto), a velocidad intermedia (162 palabras por minuto) o a gran velocidad (214 palabras por minuto). El orador respaldaba la opinión que favorecía a la ley que aumentaba la edad legal para beber alcohol, sin embargo, los estudiantes estaban en desacuerdo con esta postura. La gráfica A, en la figura 4.2, muestra la relación lineal positiva entre la velocidad y el cambio de actitudes encontrada en este estudio. En una gráfica como ésta observamos dos ejes, uno horizontal y otro vertical, llamados "x" y "y", respectivamente. Los valores de la primera variable se colocan en el eje horizontal y los valores de la segunda variable se ubican en el vertical. La gráfica A muestra que una mayor velocidad del discurso está asociada con una mayor cantidad de cambio en las actitudes.

Relación lineal negativa

Las variables también pueden relacionarse de forma negativa. En una *relación lineal negativa,* el *incremento* en los valores de una variable se acompaña del *decremento* en los valores de la otra variable. Latané, Williams y Harkins (1979) esta-

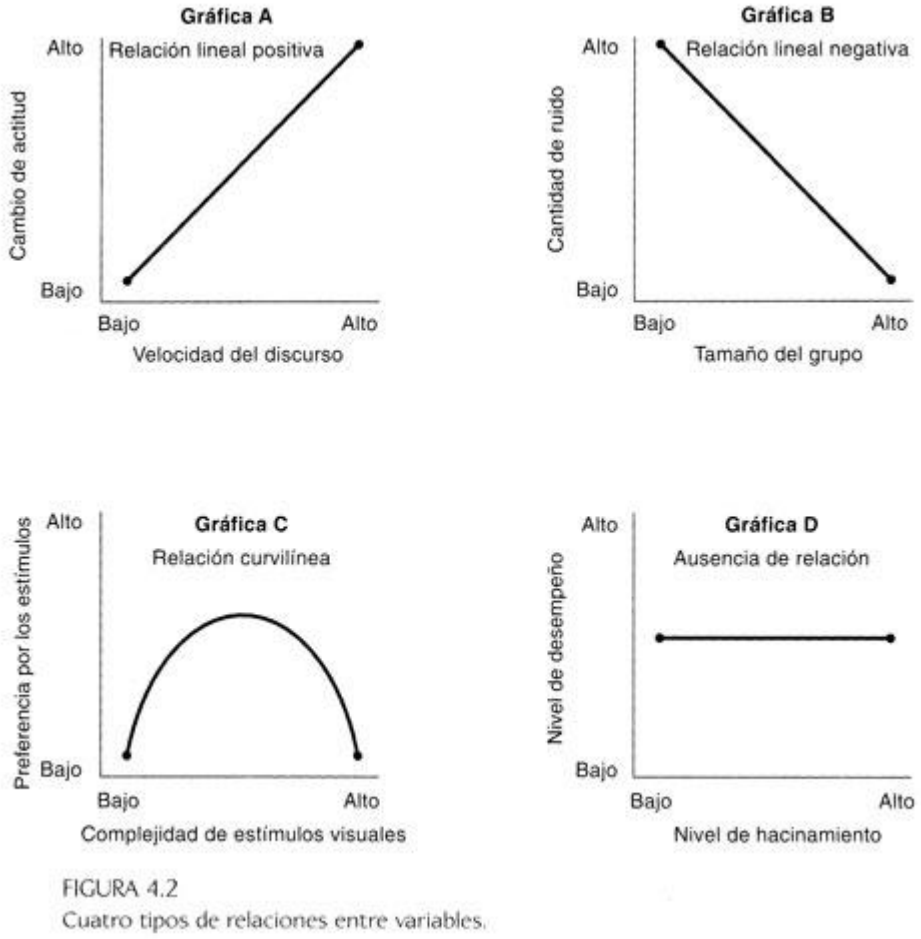


FIGURA 4.2
Cuatro tipos de relaciones entre variables.

ban intrigados por reportes que indicaban que el incremento en el número de personas que trabajan en una tarea puede, en realidad, reducir el esfuerzo y la productividad del grupo. Los investigadores diseñaron un experimento para estudiar este fenómeno, al cual denominaron "holgazanería social"; le pidieron a los participantes que aplaudieran y gritaran para hacer la mayor cantidad de ruido posible. Esto se hizo con los individuos solos, en grupos de dos, cuatro o seis personas. La gráfica B, de la figura 4.2, ilustra la relación negativa entre el número de personas en el grupo y la cantidad de ruido producida por cada una de ellas. Conforme el tamaño del grupo *aumentó*, la cantidad de ruido producido por cada persona *disminuyó*. Las dos variables están sistemáticamente relacionadas, igual que en una relación positiva, sólo que la dirección de la relación es opuesta.

Relación curvilínea

En una *relación curvilínea,* el aumento en los valores de una variable se acompaña de incrementos y decrementos en los valores de la otra variable. En otras palabras, la dirección de la relación cambia, al menos, una vez. En ocasiones, este tipo de relación se *conoce como función no monotónica.* La gráfica C, en la figura 4.2, muestra una relación curvilínea entre la complejidad de estímulos visuales y las puntuaciones de la preferencia por los estímulos. Esta relación, en particular, se conoce como relación en U invertida. El incremento en la complejidad visual se acompaña con el incremento en el gusto por el estímulo, pero únicamente hasta cierto punto. Entonces la relación se vuelve negativa y un aumento en la complejidad se ve acompañado de la *disminución* en el gusto por el estímulo (Vitz, 1966).

Ausencia de relación

Cuando no existe relación entre las dos variables, la gráfica es simplemente una línea horizontal. La gráfica D, en la figura 4.2, ilustra la relación entre el hacinamiento y el desempeño de una tarea, descubierta en un estudio llevado a cabo por Freedman, Klevansky y Ehrlich (1971). Las variables que no están relacionadas varían de forma independiente. El aumento en el número de individuos no está asociado con ningún cambio en particular en el desempeño; por consiguiente, una línea horizontal describe la ausencia de relación entre las dos variables.

Estas gráficas ilustran varios tipos de formas; casi cualquier tipo de forma puede describir la relación entre dos variables. Otras relaciones se describen por medio de formas más complejas que las que se incluyen en la figura 4.2. Por ejemplo, las relaciones lineales positiva y negativa antes descritas son ejemplos de una categoría más general de relaciones conocidas como monotónicas, porque la relación entre las variables siempre es negativa o positiva (no cambia de dirección, como en las relaciones curvilíneas o no monotónicas de la gráfica C). Un ejemplo de una función monotónica positiva que no es estrictamente lineal se muestra en la figura 4.3.

Recuerde que éstos son patrones generales. Incluso si existe, en general, una relación lineal positiva, no necesariamente significa que cada individuo que tenga un puntaje alto en una variable, también tendrá un puntaje alto en la segunda variable. Existe la posibilidad de que haya desviaciones individuales a partir del patrón general. Además de conocer el tipo general de relación entre dos variables, también es necesario conocer la fuerza de la relación; es decir, necesitamos saber el tamaño de la correlación entre las variables. En ocasiones, dos variables están fuertemente relacionadas entre sí y hay muy poca desviación del patrón general. En otras ocasiones, las dos variables no tienen una alta correlación debido a que muchos individuos se desvían del patrón general. El índice numérico de la fuerza de la relación entre variables se llama **coeficiente de correlación.** Los coeficientes de correlación son muy importantes porque necesi-

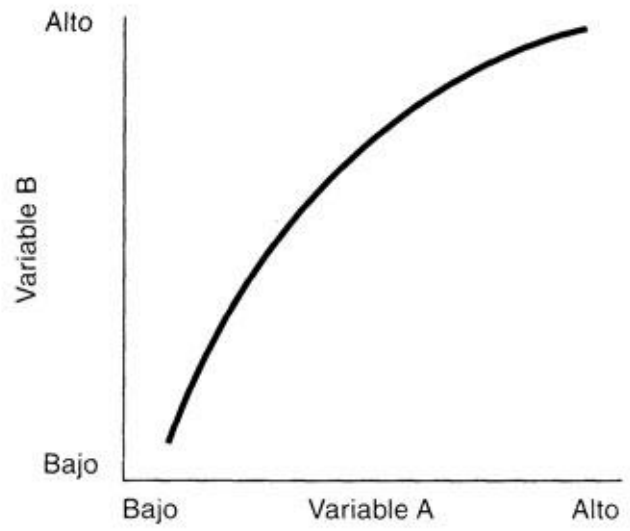


FIGURA 4.3
Función monótona positiva.

tamos saber qué tan fuerte es la relación entre las variables. Los coeficientes de correlación se estudian detalladamente en el capítulo 12.

Relaciones y reducción de la incertidumbre

Cuando detectamos una relación entre variables, reducimos la incertidumbre sobre el mundo al incrementar nuestra comprensión de las variables que estamos examinando. El término *incertidumbre* implica la aleatoriedad de los acontecimientos, los científicos se refieren a esto como *variabilidad aleatoria* o *varianza del error* de los sucesos que ocurren en el mundo. La investigación tiene la meta de reducir la variabilidad aleatoria al identificar relaciones sistemáticas entre variables.

Esto parece complejo, pero es mucho más fácil verlo en un ejemplo sencillo, para el cual las variables no tendrán propiedades cuantitativas; no describiremos *incrementos* en los valores de las variables, sino únicamente diferencias en los valores. Suponga que usted les pregunta a 200 estudiantes de su escuela si preferirían ser Mickey Mouse o Bugs Bunny (una pregunta tonta, pero útil). Ahora suponga que cien personas eligen al ratón y cien al conejo (improbable, pero continuemos). ¿Qué es lo que sabe con esta información? Lo único que sabe es que existe una variabilidad en la preferencia por estos personajes de caricaturas: algunas personas prefieren a Mickey y otras a Bugs.

Esta variabilidad se denomina varianza *aleatoria* o del *error*. Se le llama "error" únicamente porque no la comprendemos. Si usted se acerca a cualquier individuo de su escuela y trata de adivinar qué personaje prefiere, sólo tendría que hacer un pronóstico aleatorio; acertaría aproximadamente la mitad de las veces y fallaría en la misma proporción. No obstante, si pudiéramos explicar la variabilidad, ya no sería aleatorio. ¿Cómo puede reducirse la variabilidad aleatoria? La respuesta es viendo si podemos identificar las variables que están relacionadas con la preferencia de los personajes de las caricaturas.

Suponga que también pide a las personas que indiquen su género, ¿son hombres o mujeres? Ahora veamos qué sucede cuando examina si el género está

TABLA 4.1 Género y preferencia de personajes (datos hipotéticos).

		Género del participante	
		Hombres	Mujeres
Preferencia	Mickev	30	70
	Bugs	70	30
	Número de participantes	100	100

relacionado con la preferencia de los personajes de las caricaturas. La tabla 4.1 muestra un posible resultado. Observe que hay cien hombres y cien mujeres en el estudio. Sin embargo, lo importante es que 70 hombres prefirieron a Bugs y 70 mujeres eligieron a Mickey. ¿Hemos reducido la variabilidad aleatoria? Claramente lo hemos hecho. Antes de tener esta información, no hubiera habido forma de predecir si una persona fuese admiradora de Mickey o de Bugs. Ahora que *conoce* el hallazgo de la investigación, puede predecir que cualquier mujer seleccionaría a Mickey y que cualquier hombre escogería a Bugs. Ahora acertaría en el 70 por ciento de las veces, lo cual implica un gran incremento, a partir del 50 por ciento cuando todo era aleatorio.

¿Aún hay variabilidad "aleatoria"? Por supuesto, la respuesta es sí. Usted fallará aproximadamente en 30 por ciento de las veces y no sabe cuándo ocurrirá esto; por razones desconocidas, algunos hombres optarían por Mickey y algunas mujeres por Bugs. ¿Se puede reducir esta variabilidad del "error"? Esto motiva a realizar más investigación. Con otros estudios usted podría ser capaz de identificar otras variables que también están vinculadas con esta tonta elección de personajes de caricaturas.

Esta explicación revela, nuevamente, que las relaciones entre variables en pocas ocasiones son perfectas; existen hombres y mujeres que no se ajustan al patrón general.

La relación entre las variables es más fuerte cuando existe menor variabilidad aleatoria; si el 90 por ciento de los hombres elige a Bugs Bunny, la relación es mucho más fuerte (con menor incertidumbre o aleatoriedad). Si la preferencia de los hombres es de 60 por ciento, la relación es más débil.

MÉTODOS NO EXPERIMENTALES *VERSUS* EXPERIMENTALES

¿Cómo podemos determinar si las variables están relacionadas? Existen dos métodos generales para estudiar la relación entre variables: el no experimental y el experimental. Con el **método no experimental,** las relaciones se estudian haciendo observaciones o medidas de las variables de interés; es decir, la conducta se observa tal como ocurre de forma natural. Esto se puede hacer pidiendo a las personas que describan su comportamiento, por medio de observación directa,

con el registro de respuestas fisiológicas, e incluso, con el examen de diversos registros públicos, como los datos de censos. Una relación entre variables se establece cuando ambas varían en conjunto. Por ejemplo, en un estudio que describiremos en el capítulo 7, Steinberg y Dornbusch (1991) midieron el número de horas que trabajaban los estudiantes de preparatoria y relacionaron esta variable con el promedio escolar. Las dos variables cambiaron en conjunto: los estudiantes que trabajaban más horas tendían a obtener calificaciones más bajas. El segundo método para el estudio de la relación entre variables se denomina método experimental. El **método experimental** implica la manipulación y el control directo de las variables. El investigador manipula la primera variable de interés y después observa la respuesta. Por ejemplo, Loftus (1979) utilizó el método experimental cuando los participantes vieron la filmación de un accidente automovilístico y después les preguntó si habían visto "un" faro descompuesto o "el" faro descompuesto. Aquí se manipuló el método de interrogación y después se midieron las respuestas de los participantes. Con este método, las dos variables no solamente cambian en conjunto, primero se introduce una variable para ver si afecta a la segunda. Esta diferencia entre los métodos tiene implicaciones importantes.

Método no experimental

Suponga que un investigador está interesado en la relación entre el ejercicio y la ansiedad. ¿Cómo podría estudiar este tema? Si utiliza un método no experimental, el investigador podría crear definiciones operacionales para medir tanto la cantidad de ejercicio que hace la gente, como su grado de ansiedad. Podría haber una gran diversidad de formas para definir operacionalmente cualquiera de estas variables; por ejemplo, el investigador simplemente pediría a las personas que elaboren autorreportes de sus rutinas de ejercicio y de su nivel actual de ansiedad. El punto importante a recordar es que ambas variables se miden utilizando el método no experimental. Ahora suponga que el investigador reúne datos sobre el ejercicio y la ansiedad de una cantidad de individuos, y encuentra que el ejercicio está relacionado de forma negativa con la ansiedad; es decir, que quienes se ejercitan más también tienen niveles más bajos de ansiedad. Las dos variables covarían o se correlacionan entre sí: las diferencias observadas en el ejercicio están asociadas con la cantidad de ansiedad. Debido a que el método no experimental nos permite observar la covariación entre variables, otro término que con frecuencia también se utiliza es el de *método correlaciona};* con éste observamos si las variables se correlacionan o varían de manera conjunta.

El método no experimental parece ser una aproximación razonable para el estudio de las relaciones entre variables como el ejercicio y la ansiedad. Se establece una relación cuando se descubre que dos variables varían de manera conjunta; las variables covarían o se correlacionan entre sí. No obstante, este método tiene un defecto cuando planteamos preguntas sobre causa y efecto. Sabemos que dos variables están relacionadas, ¿pero qué podemos decir sobre el efecto causal de una variable sobre la otra? Existen dos problemas al hacer aseveraciones

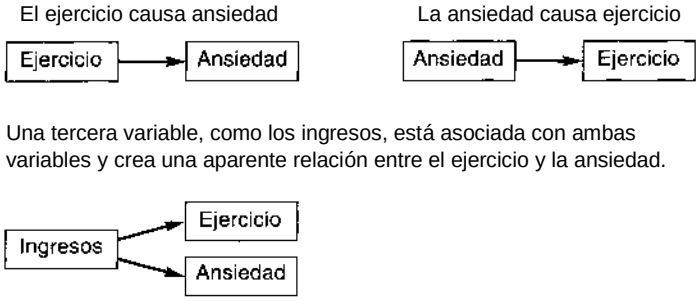


FIGURA 4.4
Posibilidades causales en un estudio no experimental.

causales cuando se utiliza el método no experimental: 1) la dirección de la causa y el efecto, y 2) el problema de la tercera variable; es decir, variables extrañas que pueden estar causando una relación observada (véase la figura 4.4, en la que se emplean flechas para escribir vínculos causales entre variables).

Dirección de causa y efecto El primer problema es el de la dirección de la causa y el efecto. Con el método no experimental es difícil determinar qué variable causa a la otra. En otras palabras, en realidad no puede decirse que el ejercicio causa una reducción de la ansiedad. A pesar de que hay razones creíbles para este patrón particular de causa y efecto, también existen razones por las que pudiese darse el patrón opuesto. Quizás el estado de ansiedad propicia que se haga ejercicio, o quizás la experiencia de una gran ansiedad interfiere con la habilidad de ejercitarse. Este es un asunto de precedencia temporal y es muy importante cuando se establecen inferencias causales. El conocimiento de la dirección correcta de la causa y el efecto tiene, a su vez, implicaciones para la aplicación de los hallazgos de investigación: si el ejercicio reduce la ansiedad, entonces realizar un programa de ejercicio sería una forma razonable para disminuir la ansiedad. Sin embargo, si la ansiedad provoca que las personas dejen de ejercitarse, no es probable que con el simple hecho de obligar a alguien a hacer ejercicio, se reduzca el grado de ansiedad del individuo.

No obstante, el problema de la dirección de la causa y el efecto no es la desventaja más grave del método no experimental. Los científicos han señalado, por ejemplo, que los astrónomos pueden hacer predicciones precisas aun cuando no pueden manipular las variables en un experimento. Además, la dirección de la causa y el efecto no suele ser crucial debido a que, para algunos pares de variables, el patrón causal puede operar en ambas direcciones. Por ejemplo, parece haber dos patrones causales en la relación entre las variables de similitud v atracción: 1) la similitud causa que las personas se atraigan unas a otras, y 2) la atracción causa que las personas se vuelvan más similares. En general, el problema de la tercera variable es una falla mucho más grave del método no experimental.

El problema de la tercera variable Cuando se aplica el método no experimental, existe el peligro de que no haya una relación causal directa entre las dos varia-

bles. Es probable que el ejercicio no afecte la ansiedad y que ésta no tenga un efecto causal sobre el ejercicio. En su lugar, tal vez exista una relación entre las dos variables porque alguna otra variable causa el ejercicio *y* la ansiedad. A esto se le *conoce* como *el problema de la tercera variable.* Cualquier cantidad de *terceras variables* (también llamadas **variables de confusión)** pueden ser responsables de una relación observada entre dos variables. En el ejemplo del ejercicio y la ansiedad, una tercera variable podría ser el nivel de ingresos. Quizás tener elevadas retribuciones permita a las personas disfrutar de más tiempo libre para hacer ejercicio (¡y la posibilidad de costear una membresía en un club deportivo!) y tal vez esto también disminuya la ansiedad. Si el nivel de los ingresos es la variable determinante, entonces no existe una relación causa y efecto directa entre el ejercicio y la ansiedad; la relación es causada por la tercera variable, el nivel de ingresos. La variable confusa es una explicación alternativa para la relación observada entre las variables. En el capítulo 1 comentamos que la posibilidad de plantear explicaciones alternativas para la relación observada entre dos variables, es otro factor importante cuando tratamos de inferir que una variable causa a otra.

Como se puede ver, la dirección de la causa y el efecto y las terceras variables potenciales constituyen limitaciones graves del método no experimental y, con frecuencia, no se toman en cuenta en los reportes de los resultados de investigaciones en los medios de comunicación masiva. Por ejemplo, un periódico puede reportar los resultados de un estudio no experimental que encontró una relación positiva entre la cantidad de café consumido y las posibilidades de sufrir un ataque cardiaco. Es obvio que no necesariamente existe una relación de causa y efecto entre las dos variables. Existen muchas terceras variables (por ejemplo, la ocupación, la personalidad o la predisposición genética) que podrían causar la conducta de beber café y la probabilidad de sufrir un ataque cardiaco. En suma, los resultados de este tipo de estudios son ambiguos y deben tomarse con escepticismo. Este tema se abordará nuevamente después de describir las características del método experimental.

Método experimental

El método experimental reduce esta ambigüedad en la interpretación de los resultados. Con el método experimental se manipula una variable y después se mide la otra. Si un investigador utiliza el método experimental para ver si el ejercicio reduce la ansiedad, el ejercicio sería manipulado, tal vez con dos grupos de personas, uno de los cuales hace ejercicio diariamente durante una semana y el otro se abstiene de realizarlo. Después se miden los estados de ansiedad. Suponga que las personas en el grupo de ejercicio tienen menor ansiedad que aquéllas en el grupo sin ejercicio. El investigador podría decir algo ahora acerca de la dirección de la causa y el efecto: en el experimento, el ejercicio ocurre primero en la secuencia de sucesos. Así, el grado de ansiedad no podría afectar la cantidad de ejercicio que la persona realiza.

Otra característica del método experimental es que intenta eliminar la influencia de todas las potenciales terceras variables extrañas. Esto se conoce

como control de variables extrañas. Este control suele lograrse cuando se garantiza que cada característica del ambiente, con excepción de la variable manipulada, se mantiene constante. Cualquier variable que no puede mantenerse constante se controla asegurándose de que sus efectos sean aleatorios. Por medio de la aleatorización, la influencia de cualquier variable extraña es igual en las condiciones experimentales. Ambos procedimientos se utilizan para asegurar que cualquier diferencia entre los grupos se debe a la variable manipulada.

Control experimental Con el control experimental todas las variables extrañas se mantienen constantes. Si una variable es constante, no puede ser responsable de los resultados del experimento, en otras palabras, no puede ser una variable de confusión. En el experimento de los efectos del ejercicio, el investigador querría asegurarse de que la única diferencia entre los grupos con y sin ejercicio es precisamente el ejercicio. Por ejemplo, debido que las personas en el grupo con ejercicio deben alterar su rutina diaria para ejercitarse, las personas del grupo sin ejercicio también deben cambiar su rutina. De otra manera, la disminución de la ansiedad en la situación con ejercicio podría resultar del "descanso" de la rutina diaria y no del ejercicio.

El control experimental se logra cuando se trata de manera idéntica a los participantes de todos los grupos del experimento, la única diferencia entre los grupos es la variable manipulada. En el experimento de Loftus sobre la memoria, ambos grupos observan el mismo accidente, el mismo experimentador plantea las preguntas a los dos grupos, las condiciones de luz y de otros aspectos son las mismas, etc. Cuando existe una diferencia entre los grupos en los reportes de memoria, se puede estar seguro de que la diferencia es el resultado del método de interrogación y no de alguna otra variable que no se mantuvo constante.

Aleatorización En ocasiones, es difícil mantener constantes algunas variables. Las más obvias son las características de los participantes. Considere un experimento en el que la mitad de los participantes de la investigación están en la condición con ejercicio, y la otra mitad en la condición sin ejercicio; los participantes en las dos condiciones podrían diferir en alguna tercera variable extraña, como el nivel de ingresos. Esta diferencia podría causar una relación aparente entre el ejercicio y la ansiedad. ¿Cómo puede el investigador eliminar la influencia de variables extrañas como éstas en un experimento?

El método experimental elimina la influencia de estas variables por medio de la **aleatorización.** La aleatorización asegura que la variable extraña tenga las mismas posibilidades de afectar tanto al grupo experimental como al otro grupo. Para eliminar la influencia de las características individuales, el investigador asigna a los participantes a los dos grupos, de manera aleatoria. En la práctica real, esto significa que la asignación a los grupos se determina utilizando una lista de números aleatorios. Para comprender esto, piense en los participantes del experimento formados en una línea. Conforme cada uno de ellos llega al inicio de la línea, se le asigna un número aleatorio, de forma parecida a los núme-

ros que se extraen en un juego de lotería. Si el número es impar, el individuo es asignado a uno de los grupos (por ejemplo, con ejercicio); si el número es par. se le incluye en el otro grupo (por ejemplo, sin ejercicio). Con el uso de un procedimiento de asignación aleatoria, el investigador puede confiar en que las características de los participantes en ambos grupos serán virtualmente idénticas Por ejemplo, en esta "lotería", las personas con ingresos bajos, medios y altos se distribuirán de forma equitativa en los dos grupos. De hecho, la aleatorizacíon asegura que la composición de las características individuales de los dos grupos será virtualmente idéntica en todos los aspectos. Esta posibilidad de asignar aleatoriamente a los participantes a las condiciones en el experimento constituye una diferencia importante entre los métodos experimental y no experimental.

Para hacer más concreto el concepto de la asignación aleatoria, usted podría realizar un ejercicio como el que se hizo con una caja llena de tarjetas de béisbol viejas. La caja contenía tarjetas de 50 jugadores de la Liga Americana y 50 de la Liga Nacional. Se mezclaron muy bien las tarjetas, después se seleccionaron 32 y se asignaron a "grupos" por medio de la tabla de números aleatorios del apéndice C.l. Al seleccionar cada tarjeta, se hizo la siguiente decisión: si el número aleatorio es par, el jugador se asigna al "grupo 1", y si el número es impar, se asigna al "grupo 2". Después se verificó si los dos grupos diferían en términos de la representación de cada liga. El grupo 1 tenía nueve jugadores de la Liga Americana y siete de la Nacional, mientras que el grupo 2 tenía el mismo número de jugadores de ambas ligas. ¡Ambos grupos serán virtualmente idénticos!

Cualquier otra variable que no puede mantenerse constante también se con-trola por medio de la aleatorización. Por ejemplo, muchos experimentos se llevan a cabo durante periodos de varios días o semanas, y los participantes entrar. al experimento en diversos momentos durante cada día. En tales casos, el investigador utiliza un orden aleatorio para organizar la secuencia de todas las condiciones experimentales. Este procedimiento previene una situación en la que una condición se organiza durante los primeros días del experimento, mientras la otra se estudia en días posteriores. De manera similar, los participantes en un grupo no se estudiarán únicamente durante la mañana, ni los otros únicamente durante la tarde.

El control directo y la aleatorización eliminan la influencia de cualquier variable extraña. Así, el método experimental permite una interpretación de los resultados relativamente carente de ambigüedades. Cualquier diferencia entre los grupos en la variable observada puede atribuirse, únicamente, a la influencia de la variable manipulada.

VARIABLES INDEPENDIENTE Y DEPENDIENTE

Cuando los investigadores estudian la relación entre variables, generalmente consideran que éstas tienen una conexión de causa y efecto. Es decir, una varia-

ble es considerada la "causa" y la otra, el "efecto". Por lo tanto, la credibilidad del orador se considera una causa del cambio de actitudes, mientras que el ejercicio tiene un efecto sobre la ansiedad. Los investigadores que utilizan los métodos experimental y no experimental ven a las variables de esta forma; aunque, como hemos visto, la ambigüedad con respecto a la dirección de la causa y el efecto es menor cuando se emplea el método experimental. Los científicos utilizan los términos **variable independiente y variable dependiente** para referirse a las variables que estudian. La variable que se considera la "causa" es la variable independiente, y la que representa el "efecto" es la variable dependiente.

En un experimento, la variable manipulada es la independiente y la variable que se mide es la dependiente. Una forma de recordar esta diferencia es relacionar los términos con lo que sucede a un individuo en un experimento. El investigador diseña una situación a la que se expone a los participantes, como observar un programa violento y otro no violento, o hacer ejercicio y no hacerlo. Esta situación es la variable manipulada; se le llama independiente debido a que el participante no tiene nada que ver con su ocurrencia. En el siguiente paso del experimento, el investigador mide la respuesta a la variable manipulada. El participante responde a lo que le sucedió, el investigador supone que lo que el individuo hace o dice es causado por, o depende del efecto de la variable independiente (manipulada). Entonces, la variable independiente es la que manipula el experimentador, y la variable dependiente es la conducta medida que se supone fue causada por la variable independiente.

Cuando la relación entre una variable independiente y una dependiente se registra en una gráfica, la independiente siempre se ubica sobre el eje horizontal y la dependiente sobre el vertical. Si revisa la figura 4.2, verá que se utilizó este método de representación gráfica para mostrar las cuatro relaciones. En la gráfica B, por ejemplo, la variable independiente, el "tamaño del grupo", está ubicada sobre el eje horizontal; la variable dependiente, "cantidad de ruido", se localiza sobre el eje vertical.

Observe que algunas investigaciones se enfocan principalmente en la variable independiente, y el investigador estudia el efecto de una sola variable independiente sobre diversas conductas. Otros científicos pueden enfocarse en una variable dependiente específica y estudiar la forma en que diversas variables independientes afectan esa conducta. Para hacer más clara la distinción, considere un estudio sobre los efectos del tamaño del jurado en los resultados de un juicio. El investigador que estudie este tema podría interesarse en las repercusiones del tamaño del grupo sobre una variedad de conductas; entre otras, las decisiones del jurado y la toma de riesgos entre gerentes de empresas. Otro investigador que se interesa únicamente en las decisiones del jurado podría analizar los efectos de muchos aspectos de los juicios, como el tamaño del jurado o las instrucciones del juez sobre el comportamiento del jurado. Ambos temas conducen a investigaciones importantes.

CAUSALIDAD

En el capítulo 1 comentamos que las inferencias de causa y efecto requieren de tres elementos. Primero, debe haber una precedencia temporal: la variable causal debe ocurrir primero en el orden temporal de los sucesos y ser seguida por el efecto. El método experimental aborda el orden temporal cuando manipula primero la variable independiente *y* después observa si tiene algún efecto sobre la variable dependiente. En otras situaciones, usted puede observar el orden temporal o puede concluir de manera lógica que un orden es más posible que otro. En segundo lugar, debe existir una covariación entre las dos variables. La covariación se demuestra en el método experimental cuando los participantes en una condición experimental (por ejemplo, la condición con ejercicio) manifiestan el efecto (como la reducción de la ansiedad), mientras que los participantes en una condición de control (por ejemplo, sin ejercicio) no manifiestan el efecto. Este requisito puede ser cubierto con una comparación de "variable presente contra ausente", como en el caso del ejercicio contra el descanso; también puede cubrirse con la comparación de las cantidades de la variable, como una hora contra diez minutos de ejercicio. En tercer lugar, existe la necesidad de eliminar explicaciones alternativas creíbles para la relación observada. Una explicación alternativa se basa en la posibilidad de que una "tercera" variable confusa sea responsable de la relación que se observa. Cuando se diseña una investigación, se pone mucha atención a la eliminación de explicaciones alternativas. El método experimental inicia cuando intenta mantener estas variables constantes por medio de la asignación aleatoria *y* del control experimental Otros aspectos del control se estudiarán en capítulos posteriores. El punto principal aquí es que las inferencias sobre relaciones causales son más fuertes cuando hay menos explicaciones alternativas para las relaciones observadas.

En ocasiones imponemos requisitos aún más estrictos antes de concluir que existe una relación causal. Algunos filósofos, científicos, e incluso muchos estudiantes, argumentan que una relación de causa y efecto se prueba únicamente s: la causa es necesaria *y* suficiente para que ocurra el efecto. Suponga que usted ha determinado que leer el material para un examen está relacionado con las calificaciones del examen; los estudiantes que leen el material obtienen calificaciones más altas que aquellos que no lo hacen. Para ser necesaria, la causa debe estar presente con el fin de que ocurra el efecto. Para probar que la lectura del material es la causa de una puntuación alta en el examen, debe demostrarse que se debe leer el material para obtener un resultado satisfactorio. Para se: suficiente, la causa debe producir *siempre* el efecto. Con el propósito de proba: que leer el material es la causa, esto siempre debe dar como resultado una calificación alta en el examen.

Analicemos esta situación en términos de las condiciones necesarias y suficientes. Si hablamos de un curso en el que el examen está basado únicamente en el material del libro, probablemente sea necesario leer el libro para obtener una calificación aprobatoria. En este punto, algunos estudiantes dicen que es posible obtener una buena calificación en el examen al acudir únicamente a las

ciases, sin leer el libro. Esto puede ser verdad en algunos casos, pero es adecuado limitar las condiciones cuando se habla de relaciones de causa y efecto.

¿Es suficiente leer el material para obtener una buena calificación en el examen? Es decir, ¿leer el material siempre da como resultado una alta calificación? Quizás usted esté pensando en las muchas ocasiones en que leyó el material, pero *no* obtuvo una alta calificación. Leer el material no es una causa suficiente; se tienen mayores posibilidades de aprender el material cuando se pone atención, cuando se relaciona la información con otros conocimientos y cuando se practica la evocación del material.

Los requisitos "necesario *y* suficiente" para establecer la causa son poco comunes en la psicología. Siempre que los psicólogos aseveran que existe una causa necesaria y suficiente para una conducta, la investigación pronto revela que sencillamente no es así. Por ejemplo, un grupo de psicólogos aseveró alguna vez que "la frustración causa agresión"; siempre que haya frustración, resulta la agresión, y siempre que hay agresión, la frustración debe ser su causa preceden-re. Se demostró que esta afirmación es inexacta. La frustración puede conducir a la agresión, pero también existe la posibilidad de otras respuestas, como el aislamiento pasivo o un mayor esfuerzo para superar la frustración. Además, la agresión puede resultar de la frustración; pero otros acontecimientos pueden producirla también, entre ellos el dolor, los insultos o un ataque directo.

Los científicos del comportamiento no se preocupan demasiado por el cenia de la causa y el efecto últimos. En su lugar, están más interesados en describir cuidadosamente la conducta, en estudiar la forma en que las variables se afectan entre sí y las teorías que explican el comportamiento. El consenso general es que existen pocas causas "necesarias y suficientes" interesantes del comportamiento. En su lugar, la investigación de diversas variables eventualmente conduce a la comprensión de una "red causal" completa, en la que una cantidad de variables está involucrada en patrones complejos de causa y efecto. Este libro no trata estas difíciles cuestiones, sino que examina los métodos utilizados para estudiar el comportamiento.

ELECCIÓN DE UN MÉTODO: VENTAJAS DE DIVERSOS MÉTODOS

Se han destacado las ventajas del método experimental para el estudio de relaciones entre variables. Sin embargo, *existen* desventajas en los experimentos y muchas razones para utilizar otros métodos y no el experimento. Examinemos algunos de los temas que surgen cuando se elige un método.

Artificialidad de los experimentos

En un experimento se manipúlala variable independiente, generalmente dentro de los confines cuidadosamente controlados de un laboratorio. Este procedimiento permite hacer inferencias relativamente menos ambiguas con respecto a

causas y efectos, y reduce la posibilidad de que variables extrañas puedan afectar los resultados. La experimentación de laboratorio es una forma sumamente valiosa para estudiar muchos problemas. No obstante, el alto grado de control y el ambiente del laboratorio en ocasiones crean una atmósfera artificial que puede limitar las preguntas que se abordan o la generalidad de los resultados. Por esta razón, los investigadores pueden decidir el uso de métodos no experimentales.

Otra opción es intentar realizar un experimento en un ambiente de campo. En un **experimento de campo** la variable independiente se manipula en un ambiente natural. Como en cualquier experimento, el investigador trata de controlar las variables extrañas por medio de la aleatorización o el control experimental. Como ejemplo de un experimento de campo, considere el estudio de Langer y Rodin (1976), referente a los efectos de otorgar a los residentes de asilos de ancianos mayor control sobre las decisiones que afectan sus vidas. A un grupo de residentes se le otorgó una gran responsabilidad para tomar decisiones con respecto a la operación del asilo; a un segundo grupo se le hizo sentir que el personal sería responsable de su cuidado y de sus necesidades. Los investigadores midieron variables dependientes como el nivel de actividad y la felicidad de los residentes. Los resultados mostraron que las personas en el grupo de mayor responsabilidad eran más activas y felices. En un estudio de seguimiento, estos residentes incluso manifestaron una gran mejoría en su salud física (Rodin y Langer, 1977).

Muchos otros experimentos de campo se llevan a cabo en espacios públicos como las esquinas de las calles, tiendas de autoservicio y estacionamientos Ruback y Juieng (1997) midieron el tiempo que tardaban en dejar su lugar los conductores en un estacionamiento, bajo dos condiciones: 1) cuando otro automóvil estaba esperando el lugar, o 2) cuando no había otro vehículo presente Como era de esperarse, los conductores tardaban más en irse cuando un automóvil estaba esperando el lugar. ¡Aparentemente, la motivación de proteger un territorio temporal es más fuerte que la motivación de irse lo más pronto posible! La ventaja del experimento de campo es que la variable independiente se investiga en un contexto natural. La desventaja es que el investigador pierde k posibilidad de controlar, de forma directa, muchos aspectos de la situación. El experimento de laboratorio permite a los científicos mantener, con mayor facilidad, constantes a las variables extrañas, eliminando así su influencia en los resultados del experimento. Por supuesto, es precisamente este control el que conduce a la artificialidad de la investigación de laboratorio. Afortunadamente, cuando los investigadores han realizado experimentos en ambientes de laboratorio y de campo, los resultados han sido muy similares (Anderson, Lindsay v Bushman, 1999).

Consideraciones éticas y prácticas

En ocasiones, el método experimental no es una alternativa factible, ya que la experimentación sería poco ética o impracticable. Por ejemplo, sería inútil ma-

nipular las prácticas de crianza infantil con el método experimental. Incluso, si fuese posible asignar aleatoriamente a los padres a dos condiciones de crianza infantil, como el retiro del amor contra los tipos de castigo físico, la manipulación sería poco ética. En lugar de manipular variables como las técnicas de crianza infantil, los investigadores suelen estudiarlas tal como ocurren en ambientes naturales. Muchas áreas de investigación importantes presentan problemas similares, por ejemplo, los estudios sobre los efectos del alcoholismo, el divorcio y sus consecuencias o las repercusiones del empleo de la madre sobre los hijos. Problemas como éstos deben ser estudiados y, generalmente, las únicas técnicas posibles son de tipo no experimental.

Cuando se estudian variables como éstas, con frecuencia se clasifica a las personas en grupos con base en sus experiencias. Al estudiar el empleo materno, un grupo podría consistir en individuos cuyas madres trabajan fuera del hogar y otro grupo de individuos cuyas madres no trabajan. Este se *conoce* también como un diseño *ex post facto. Ex post facto* significa "después del hecho". El término se acuñó para describir investigaciones en las que los grupos se forman con base en alguna diferencia real y no por la asignación aleatoria, como en un experimento. Es extremadamente importante estudiar estas diferencias. Sin embargo, es necesario reconocer que se trata de investigación no experimental, ya que no hay una asignación aleatoria de los grupos.

Variables de los participantes

Las variables de los participantes (también llamadas *variables de los sujetos* y *variables atributivas)* son las características de los individuos como la edad, el género, el grupo étnico, la nacionalidad, el orden de nacimiento, la personalidad o el estado civil. Por definición, estas variables son no experimentales y por lo tanto deben medirse. Por ejemplo, para estudiar una característica de la personalidad como la extroversión, podría pedírsele a las personas que completaran una prueba de personalidad que esté diseñada para medir esta variable. Este tipo de variables pueden estudiarse en experimentos junto con variables independientes manipuladas (véase el capítulo 10).

Descripción del comportamiento

Una de las principales metas de la ciencia es proporcionar descripciones precisas de acontecimientos. En consecuencia, la meta de muchas investigaciones es describir el comportamiento; los aspectos que abordan los experimentos no son "relevantes para las principales metas de la investigación. Un ejemplo clásico de investigación descriptiva en psicología proviene del trabajo de Jean Piaget, quien observó cuidadosamente el comportamiento de sus propios hijos conforme maduraban, y describió con detalle los cambios en sus formas de pensamiento sobre su ambiente y sus respuestas ante éste (Piaget, 1952). Las descripciones de Piaget y las interpretaciones de sus observaciones, dieron como resultado una importante teoría del desarrollo cognoscitivo que incrementó enormemente

nuestra comprensión de este tema. La teoría de Piaget tuvo un gran impacto en la psicología, el cual continúa hoy día (Flavell, 1996).

Predicciones exitosas del comportamiento futuro

En muchas situaciones de la vida real, un objetivo importante es hacer predicciones exitosas del comportamiento futuro de una persona; por ejemplo, el éxito escolar, la habilidad para aprender un nuevo empleo o los probables intereses en diversos campos en la universidad. En tales circunstancias, tal vez no exista la necesidad de preocuparse por aspectos de causa y efecto. Es posible diseñar medidas que incrementen la precisión de las predicciones del comportamiento futuro. Los consejeros escolares pueden aplicar pruebas para decidir si los estudiantes deben estar en programas "enriquecidos"; los empleadores pueden probar a los solicitantes para determinar si deben ser contratados; los estudiantes universitarios pueden responder pruebas que los ayuden a decidir qué carrera estudiar. Estos tipos de medidas pueden conducir a mejores decisiones para muchas personas. Cuando los investigadores elaboran medidas diseñadas para predecir el comportamiento futuro, deben realizar investigaciones para demostrar que la medida se relaciona con el comportamiento en cuestión. Este tipo de investigación se estudiará en el capítulo 5.

Ventajas de diversos métodos

Quizás lo más importante es que la plena comprensión de cualquier fenómeno requiere de un estudio con el uso de diversos métodos, tanto experimentales como no experimentales. Ningún método es perfecto y ningún estudio es definitivo, Para ilustrar esto, considere una hipótesis desarrollada por Frank y Gi-lovich (1988). Ellos estaban intrigados por la observación de que el color negro representa lo diabólico y la muerte en muchas culturas, y se preguntaron si esto tiene una influencia sobre nuestro comportamiento. Observaron que varios equipos deportivos profesionales de la Liga Nacional de Fútbol Americano (NFL) y de la Liga Nacional de Hockey (NHL) utilizan uniformes negros, por lo cual elaboraron la hipótesis de que estos equipos podrían ser más agresivos que otros en las ligas.

Primero necesitaban una definición operacional de uniformes "negros" y "no negros"; decidieron que un uniforme negro es aquél en el que 50 por ciento o más de su hechura es de ese color. Con esta definición, cinco equipos de la NFL y cinco de la NHL tenían uniformes negros. Primero pidieron a varias personas que no tenían conocimiento sobre la NFL o la NHL que vieran el uniforme de cada equipo, y después calificaran a los equipos con adjetivos "malévolos" como "malo" y "agresivo". En general, los equipos *con* uniformes negros fueron percibidos como los más agresivos. Después compararon las yardas penalizadas de equipos negros y no negros de la NFL y los minutos de castigo de los equipos de la NHL. En ambos casos, los equipos negros recibieron más sanciones. ¿Pero existe un patrón causal? Frank y Gilovich descubrieron que dos equipos

de la NHL habían cambiado los uniformes al negro, por lo cual compararon los minutos de castigo antes y después del cambio; de forma consistente con la hipótesis, las penalizaciones se incrementaron para ambos equipos. También registraron los minutos de castigo de un tercer equipo que cambió el uniforme de un color que no era negro a otro color que tampoco lo era, y no encontraron ningún cambio en los minutos de sanción. Observe que ninguno de estos estudios utilizó el método experimental. En un experimento para probar la hipótesis de que las personas perciben como más agresivos a los equipos de uniforme negro, diversos estudiantes observaron filmaciones de dos juegos de fútbol en el que la defensa estaba utilizando el blanco o el negro. Ambos juegos incluían un acto agresivo realizado por la defensa. En estos encuentros, los estudiantes penalizaron más al equipo con uniforme negro que al equipo sin uniforme negro. En un experimento final, realizado para ver si pertenecer a un equipo con uniforme negro aumenta la agresividad, se llevó a unas personas al laboratorio en grupos de tres. Se les dijo que eran un "equipo" que competiría con otro. A todos los miembros se les dio ropa blanca o negra para la competencia; se les pidió que eligieran los juegos que querían jugar para la competencia. Algunos de los juegos eran agresivos ("duelo con pistola de dardos") y otros no ("competencia de golf). Como era de esperarse, los equipos con uniforme negro eligieron los juegos más agresivos.

El punto importante aquí es que ningún estudio constituye una prueba de hipótesis perfecta. Sin embargo, cuando diversos estudios, con el uso de variados métodos, llegan a la misma conclusión, nuestra confianza en los hallazgos y nuestra comprensión del fenómeno se incrementa de forma importante.

EVALUACIÓN DE LA INVESTIGACIÓN: TRES TIPOS DE VALIDEZ

La *validez* se refiere a la "verdad" y a la representación precisa de la información. La investigación puede describirse y evaluarse en términos de tres tipos de validez: la validez de constructo, la validez interna y la validez externa. Cada uno nos ofrece una perspectiva diferente sobre cualquier investigación en particular.

Validez de constructo

La **validez de constructo** se refiere a la adecuación de la definición operacional de las variables: ¿la definición operacional de una variable refleja realmente el verdadero significado teórico de dicha variable? Muchas variables son "constructos" abstractos, como la ansiedad social, la credibilidad del orador y la holgazanería social. Se debe crear una medida de la ansiedad social para evaluar un constructo teórico; es decir, una definición operacional de la variable. La medida tiene validez de constructo si mide el constructo de la ansiedad social y no alguna otra variable como el dominio. De forma similar, la manipulación de la credibilidad de un orador, con oradores de baja y alta credibilidad, debe ma-

nipular la credibilidad y no otra cosa, como el atractivo. Puesto que las variables pueden medirse y manipularse de diversas formas, nunca existe una definición operacional perfecta para una variable. Con el tiempo, otros investigadores utilizarán diversos métodos para definir operacionalmente cualquier variable. Este tema se estudiará con mayor profundidad en próximos capítulos.

Validez interna

La **validez interna** se refiere a la posibilidad de sacar conclusiones sobre relaciones causales, a partir de nuestros datos. Un estudio tiene una alta validez interna cuando se pueden hacer inferencias firmes acerca de que una variable causó cambios en la otra variable. Hemos visto que se pueden hacer inferencias causales firmes con mayor facilidad cuando se utiliza el método experimental. La validez interna se incrementa cuando las consideraciones de causa y efecto, que se discutieron anteriormente, pueden aplicarse a la investigación.

Validez externa

Finalmente, la **validez externa** de un estudio es la magnitud en que los resultados pueden generalizarse a otras poblaciones y ambientes. ¿Pueden replicarse los resultados con otras definiciones operacionales de las variables, con distintos participantes y en otros ambientes? En este capítulo se abordaron los aspectos de la artificialidad de los experimentos de laboratorio, se trata de un asunto de validez externa. Note que la meta de una alta validez interna en ocasiones puede entrar en conflicto con la meta de la validez externa. Los experimentos de campo representan una forma en que los investigadores tratan de incrementar la validez externa de sus estudios. El tema de la validez externa es complejo y se abordará con mayor detalle en el capítulo 14.

Hasta este punto, usted puede estar preguntándose cómo es que los investigadores seleccionan una metodología para estudiar problemas. Se dispone de una variedad de métodos, cada uno con ventajas y desventajas. Los investigadores seleccionan el método que les permite abordar mejor las preguntas que desean responder. Ningún método es inherentemente superior a otro. Más bien, la elección del método se hace después de considerar el problema de investigación, la ética, las limitaciones por los costos y el tiempo, y los aspectos asociados con los tres tipos de validez. En lo que resta de este libro, se discutirán muchos métodos específicos, todos los cuales son útiles bajo distintas circunstancias. De hecho, todos son necesarios para comprender la gran variedad de conductas que son de interés para los científicos del comportamiento. La comprensión total de cualquier problema o tema requiere de investigación que utilice una variedad de aproximaciones metodológicas.

Términos de estudio

Aleatorización Causa necesaria Causa suficiente Coeficiente de correlación Confuso Control experimental Definición operacional Experimento de campo Método experimental Método no experimental (método correlacional) Problema de la tercera variable Relación curvilínea Relación lineal negativa Relación lineal positiva Validez de constructo Validez externa Validez interna Variable Variable de confusión Variable de los participantes (sujetos) Variable de respuesta Variable dependiente Variable independiente Variable mediadora Variable situacional

Preguntas de repaso

1. ¿Qué es una variable? Enumere al menos *cinco* variables diferentes y después especifique los niveles de cada una.
2. Describa la "definición operacional" de una variable. Dé al menos dos definiciones operacionales de las variables que pensó en la pregunta anterior.
3. Distinga entre la relación lineal positiva, la lineal negativa y la curvilínea.
4. ¿Cuál es la diferencia entre el método no experimental y el método experimental?
5. ¿Cuál es la diferencia entre una variable independiente y una variable dependiente?
6. Diga cuál es la diferencia entre un experimento de laboratorio y uno de campo.
7. ¿A qué se refieren el problema de la dirección de la causa y el efecto, y el problema de la tercera variable?
8. ¿De qué forma influyen el control experimental directo y la aleatorización en los posibles efectos de variables extrañas?
9. ¿Cuáles son algunas de las razones para utilizar el método no experimental en el estudio de relaciones entre variables?
10. Describa los tres elementos para inferir causalidad.
11. ¿A qué se refiere una causa "necesaria y suficiente"?

Actividades

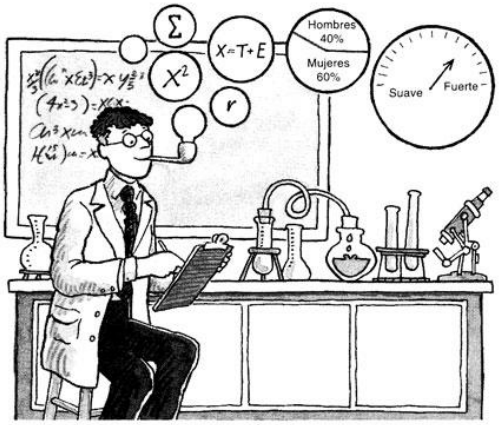
1. Los hombres y las mujeres pueden diferir en la forma en que ayudan a otras personas. Por ejemplo, existen mayores posibilidades de que los hombres ayuden a una persona que tenga un problema con su automóvil, y las mujeres tienen mayores posibilidades de llevarle comida a un amigo enfermo. Elabore dos definiciones operacionales para el concepto de la conducta de ayuda, uno que destaque el "estilo masculino" y el otro, el "estilo femenino". ¿De qué forma el uso de uno u otro podría conducir a conclusiones distintas, a partir de resultados experimentales, con relación a quién ayuda más, los hombres o las mujeres? ¿Qué indica esto acerca de la importancia de las definiciones operacionales?
2. Usted observa que los compañeros de clase que tienen buenas calificaciones tienden a sentarse al frente del salón de clases, mientras que quienes obtienen malas calificaciones suelen sentarse en la parte de atrás. ¿Cuáles son tres posibles relaciones de causa y efecto en esta observación de tipo no experimental?
3. Considere la hipótesis de que el estrés en el trabajo causa conflictos familiares en el hogar.
   a) ¿Qué tipo de relación se propone (por ejemplo, lineal positiva, lineal negativa)?
   b) Haga una gráfica de la relación propuesta.
   c) Identifique la variable independiente y la variable dependiente en el planteamiento de la hipótesis.
   d) ¿Cómo investigaría la hipótesis por medio del método experimental?
   é) ¿Cómo investigaría la hipótesis utilizando el método no experimental (reconociendo los problemas para determinar la causa y el efecto)?
   f) ¿Qué factores consideraría para decidir si emplea el método experimental o el no experimental para estudiar la relación entre el estrés del trabajo y el conflicto familiar?
4. Identifique las variables independiente y dependiente en las siguientes descripciones de experimentos:
   a) Varios estudiantes vieron una caricatura solos o con otras personas, y después evaluaron qué tan graciosa les pareció.
   b) Se aplicó una prueba de comprensión a diversos alumnos después de haber estudiado el material de un libro de texto en silencio o con el televisor encendido.
   c) Se les dijo a algunos maestros de primaria que los padres de un niño tenían un título universitario, y a otros maestros se les dijo que los padres del niño no habían terminado la preparatoria; después evaluaron el potencial académico del niño.

d) Los trabajadores de una empresa fueron asignados a una de las siguientes condiciones: un grupo completó un programa de entrenamiento para el manejo del estrés; otro grupo de trabajadores no participó en el entrenamiento. Durante los dos meses siguientes se examinó el número de días que los trabajadores no acudían al trabajo por enfermedad.

5. Hace algunos años los periódicos reportaron el hallazgo de que los estadounidenses que toman una copa de vino al día son más saludables que aquellos que no toman vino (o que toman mucho vino u otra bebida alcohólica). ¿Cuáles son algunas explicaciones alternativas posibles para este hallazgo, es decir, que otras variables, además del vino, podrían explicar el descubrimiento? (Pista: ¿Qué tipo de personas en Estados Unidos tienen mayores posibilidades de tomar una copa de vino con la cena?)

2. Las limitaciones de la investigación no experimental llamaron drásticamente la atención del público por los resultados de un experimento sobre los efectos de la terapia de reemplazo hormonal posterior a la menopausia (parte de un gran estudio conocido como Women's Health Initiative). En la investigación médica, a un experimento se le llama ensayo clínico. En el ensayo clínico se asignó aleatoriamente a los participantes para recibir la terapia de sustitución hormonal o un placebo (sin hormonas). La terapia de reemplazo hormonal consistió de estrógenos y progestina. En 2002, los investigadores concluyeron que las mujeres que tomaron la terapia de reemplazo hormonal presentaron una incidencia más alta de enfermedades cardiacas que las mujeres en la condición del placebo (sin hormonas). En ese punto detuvieron el experimento e informaron a las participantes y al público que debían hablar con sus médicos acerca de la conveniencia de esta terapia. El hallazgo se contrapuso de forma dramática con los resultados de una investigación no experimental, en la cual las mujeres que tomaban hormonas tuvieron una incidencia menor de enfermedades cardiacas; en estos estudios los investigadores compararon mujeres que ya tomaban las hormonas con otras que no las tomaban. ¿Por qué cree que los resultados fueron distintos en la investigación experimental y en la no experimental?

5



Conceptos sobre medición

Confiabilidad de la medida
Confiabilidad de *test-retest* Confiabilidad de consistencia interna Confiabilidad entre jueces Confiabilidad y precisión de las mediciones

Validez de constructo de las mediciones
Indicadores de la validez de constructo
Validez predictiva Validez concurrente
Validez convergente Validez discriminante

Reactividad de las mediciones

Variables y escalas de medición
Escalas nominales
Escalas ordinales
Escalas de intervalo y de razón
La importancia de las escalas de medición

Términos de estudio

Preguntas de repaso

Actividades

Aprendemos sobre el comportamiento por medio de mediciones cuidadosas. Como estudiamos en el capítulo 4, la conducta puede medirse de muchas formas. La estrategia de medición más común es preguntar a las personas acerca de sí mismas: ¿cuántas veces discutió con su esposo(a) la semana pasada?, ¿cómo calificaría su felicidad, en general?, ¿qué tanto le gustó su compañero en este experimento? Por supuesto, usted también puede observar el comportamiento de forma directa. ¿Cuántos errores cometió alguien en una tarea?, ¿le cambiarán un billete las personas a quienes se acerca en una tienda de autoservicio?, ¿cuántas veces sonrió una persona durante una entrevista? También pueden medirse las respuestas fisiológicas y neurológicas. ¿Qué tanto cambió la frecuencia cardiaca mientras se resolvían los problemas?, ¿se incrementó la tensión muscular durante la entrevista? Existe una infinidad de conductas fascinantes que pueden estudiarse. Describiremos varios métodos de medición de variables en diversos puntos de los siguientes capítulos. Sin embargo, en este capítulo exploramos los aspectos técnicos de la medición. Necesitamos considerar la confiabilidad, la validez y la reactividad de las mediciones. También abordaremos las escalas de medición.

CONFIABILIDAD DE LA MEDIDA

La **confiabilidad** se refiere a la consistencia o estabilidad de una medida del comportamiento. La definición cotidiana de la confiabilidad se acerca bastante a la definición científica. Por ejemplo, se podría decir que la profesora Fuentes es "confiable" porque inicia su clase exactamente a las 10:00 de la mañana, diariamente; en contraste, a la profesora Fine se le podría considerar "poco confiable" debido a que, aunque en ocasiones inicia su clase exactamente a la hora, cualquier día puede llegar entre las 10:00 y las 10:20 de la mañana.

De forma similar, una medida confiable de una variable psicológica, como la inteligencia, puede producir el mismo resultado cada vez que se administra la prueba de inteligencia a la misma persona. El examen sería poco confiable si evalúa a la misma persona en una primera semana como promedio, como deficiente la siguiente semana y brillante en la tercera semana. En términos sencillos, una medida confiable no fluctúa de una lectura a la siguiente. Si la medida fluctúa, existe un error en el instrumento de medición.

Una manera más formal de comprender la confiabilidad implica el uso de los conceptos de puntaje verdadero y error de medición. Se puede pensar que cualquier medida que se haga tiene estos componentes: 1) un **puntaje verdadero,** que es el puntaje real de la variable, y 2) un **error de medición.** Una medida poco confiable de inteligencia contiene una cantidad cosiderable de error de medición, por lo cual no proporciona un indicador preciso de la verdadera inteligencia del individuo. En contraste, una medida confiable de la inteligencia —que contenga poco error de medición— producirá un puntaje idéntico (o casi idéntico), cada vez que se mida al mismo individuo.

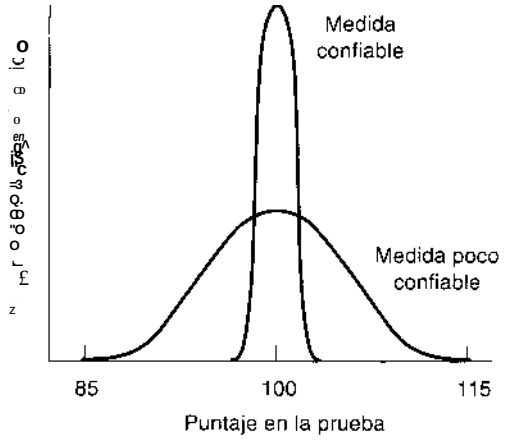


FIGURA 5.1
Comparación de ¡os datos de una medida confiable y de otra poco confiable.

Para ilustrar un poco más el concepto de confiabilidad, imagínese que usted *conoce* a alguien cuyo "verdadero" puntaje de inteligencia es de 100. Ahora suponga que usted le administra una prueba poco confiable de inteligencia cada semana durante un año. Una vez que pasa el año, calcula el puntaje promedio del individuo en la prueba, con base en los 52 puntajes obtenidos. Ahora suponga de nuevo, que prueba a otro amigo que también tiene una puntuación verdadera de inteligencia de 100; sin embargo, en esta ocasión usted administra una prueba con alta confiabilidad. Nuevamente, calcula el puntaje promedio. ¿Cómo serían sus datos? Los datos típicos se muestran en la figura 5.1. En cada caso, el puntaje promedio es de 100. No obstante, los puntajes de la prueba poco confiable oscilan entre 85 y 115, mientras que los puntajes de la prueba confiable van de 97 a 103. El *error de medición* de la prueba poco confiable se revela en la mayor variabilidad que mostró la persona que respondió esta prueba.

Cuando se realiza una investigación, sólo se puede medir a cada persona una vez; no puede hacer la medición 50 o 100 veces para descubrir un puntaje verdadero. Por lo tanto, es muy importante que se utilice una medida confiable. La administración única de la medida debe reflejar con mucha precisión el puntaje verdadero de la persona.

La importancia de la confiabilidad es obvia. Una medida de longitud poco confiable sería inútil para construir una mesa; una medida poco confiable de una variable, como la inteligencia, es también inútil para el estudio de ésta. Los investigadores no pueden utilizar medidas poco confiables para estudiar de forma sistemática variables o la relación que hay entre éstas. Tratar de estudiar el comportamiento por medio de medidas poco confiables representa una pérdida de tiempo, ya que los resultados serán inestables y no podrán replicarse.

Existen mayores posibilidades de lograr confiabilidad cuando los investigadores utilizan procedimientos de medición cuidadosos. En algunas áreas de investigación esto podría implicar observadores cuidadosamente entrenados que registren el comportamiento; en otras áreas, podría ser necesario poner gran

atención a la forma en que se plantean las preguntas o a la manera en que se colocan los electrodos en el cuerpo para medir reacciones fisiológicas.

¿Cómo podemos evaluar la confiabilidad? No podemos observar de manera directa el puntaje verdadero ni los componentes de error de un puntaje real de la medida. Sin embargo, podemos evaluar la estabilidad de las medidas utilizando coeficientes de correlación. Recuerde que en el capítulo 4 estudiamos que un coeficiente de correlación es un número que nos indica qué tan fuerte es la relación entre dos variables. Existen varias formas para calcular los coeficientes de correlación; el coeficiente de correlación más común para la evaluación de la confiabilidad es el **coeficiente de correlación producto-momento de Pearson.** Este coeficiente de correlación (simbolizado por r) varía desde 0.00 hasta +1.00 v desde 0.00 hasta -1.00. Una correlación de 0.00 indica que las dos variables no tienen ninguna relación. Mientras más se acerque una correlación a 1.00, ya sea -1.00 o -1.00, más fuerte será la relación. Los signos positivo y negativo proporcionan información sobre la dirección de la relación. Cuando el coeficiente de correlación es positivo (un signo de "más"), existe una relación lineal positiva, los puntajes altos en una variable están asociados con puntajes altos en la segunda variable. Una relación lineal negativa se indica por un signo de "menos", los puntajes altos en una variable están asociadas con los puntajes bajos en la segunda variable. El coeficiente de correlación de Pearson se estudiará con mayor profundidad en el capítulo 12.

Para evaluar la confiabilidad de una medida, necesitamos obtener al menos dos puntajes de dicha medida, de muchos individuos. Si la medida es confiable, los dos puntajes deben ser muy similares; un coeficiente de correlación de Pearson que relacione los dos puntajes debe indicar una correlación positiva alta. Cuando lea acerca de la confiabilidad, generalmente la correlación se denominará *coeficiente de confiabilidad.* Examinemos métodos específicos para evaluar la confiabilidad.

Confiabilidad de test-retest

La **confiabilidad de *test-retest*** se evalúa midiendo a los mismos individuos en dos momentos diferentes. Por ejemplo, se podría calificar la confiabilidad de una prueba de inteligencia midiendo a un grupo de personas un día y, nuevamente, una semana después. Entonces tendríamos dos puntuaciones para cada persona, y calcularíamos el coeficiente de correlación para determinar la relación entre los puntajes del *test y* los puntajes del *retest.* Recuerde que una alta confiabilidad se indica por un elevado coeficiente de correlación, el cual muestra que los dos puntajes son muy similares. Si muchas personas tienen puntajes muy similares, concluiremos que la medida refleja puntajes verdaderos, más que el error de medición. Es difícil decir qué tan alta debe ser la correlación para poder considerar que la medida es confiable, sin embargo para la mayoría de las medidas el coeficiente de confiabilidad probablemente deba ser de al menos 0.80.

Puesto que la confiabilidad de *test-retest* requiere que se administre la misma prueba dos *veces,* la correlación podría ser artificialmente alta, ya que los índivi-

duos recuerdan cómo respondieron la primera vez. En ocasiones se utilizan formas alternativas de confiabilidad para evitar este problema. Estas incluyen la administración de dos formas distintas de la misma prueba a los mismos individuos en dos momentos diferentes.

La inteligencia es una variable que, podría esperarse que permanece relativamente constante a lo largo del tiempo. Por consiguiente, esperamos que la confiabilidad de *test-retest* de la inteligencia sea muy alta. No obstante, se espera que algunas variables cambien de un periodo de prueba al siguiente. Por ejemplo. una escala diseñada para medir el estado de ánimo de una persona es una medida que podría cambiar fácilmente de un periodo de prueba a otro, y la confiabilidad de *test-retest* quizás no sea adecuada. Desde un punto de vista más práctico, en ocasiones es difícil obtener dos medidas de la misma persona en dos momentos diferentes. Para resolver esto, los investigadores han diseñado métodos para evaluar la confiabilidad sin la necesidad de emplear dos pruebas separadas.

Confiabilidad de consistencia interna

Es posible evaluar la confiabilidad midiendo a los individuos en un solo momento. Podemos hacer esto debido a que la mayoría de las medidas psicológicas están constituidas por un número de preguntas diferentes, llamadas reactivos. Una prueba de inteligencia puede tener 100 reactivos, una medida de la extro-versión 15, y un examen de opción múltiple para una clase podría tener 50. El puntaje de la prueba de una persona podría basarse en el total de sus respuestas en todos los reactivos. En una clase, un examen consiste en un número de preguntas sobre el material, y el puntaje total es la cantidad de respuestas correctas. En una medida de la extroversión se podría pedir a las personas que manifiesten su acuerdo o desacuerdo con reactivos tales como "disfruto la estimulación de una fiesta animada". El puntaje de extroversión de un individuo se obtiene a partir del número total de reactivos en los que estuvo de acuerdo.

La **confiabilidad de consistencia interna** se evalúa utilizando respuestas en un solo momento. Puesto que todos los reactivos miden la misma variable, deben producir resultados similares o consistentes. Un indicador de consistencia interna es la **confiabilidad por mitades;** se trata de la correlación del puntaje total de un individuo en la mitad de la prueba, con el puntaje total en la otra mitad. Las dos mitades se crean dividiendo aleatoriamente los reactivos en dos partes. El cálculo real del coeficiente de confiabilidad por mitades es un poco mas complicado, ya que la medida final incluye reactivos de ambas mitades. Así, la medida combinada tiene más reactivos y es más confiable que cualquiera de las dos mitades de manera aislada. Debe tomarse en cuenta este hecho cuando se calcula el coeficiente de confiabilidad; consulte un libro sobre medición psicológica para estudiar los procedimientos técnicos, si necesita realizar los cálculos.

Otro indicador de confiabilidad de consistencia interna es la llamada *alfa* **de Cronbach.** Aquí, el investigador calcula la correlación de cada reactivo con cada uno de los otros reactivos; resulta una gran cantidad de coeficientes de correlación, ¡por lo cual tendrá que hacerlo en una computadora! El valor de *alfa*

es el promedio de todos los coeficientes de correlación. También se puede examinar la correlación del puntaje de cada reactivo con el puntaje total, basado en todos los reactivos. Este tipo de correlaciones reactivo-total y el *alfa* de Cronbach brindan mucha información, ya que proporcionan datos sobre cada reactivo individual. Los reactivos que no se correlacionan con otros pueden eliminarse de la medida para aumentar la confiabilidad.

Confiabilidad entre jueces

En ciertas investigaciones, algunos evaluadores observan las conductas y hacen puntajes o juicios. Para hacer esto, un evaluador o juez sigue instrucciones para hacer juicios sobre las conductas; por ejemplo, evaluando si el comportamiento de un niño en un parque es agresivo, y qué tanto lo es. Usted puede pedir a un evaluador que haga juicios sobre la agresión, pero las observaciones de un solo evaluador podrían ser poco confiables. La solución para este problema es utilizar, al menos, dos evaluadores que observen la misma conducta. La **confiabilidad entre jueces** se refiere a la magnitud de acuerdo que hay entre los evaluadores con respecto de sus observaciones. Por lo tanto, si dos evaluadores juzgan que las conductas son agresivas, se obtiene una alta correlación entre jueces, cuando la mayoría de las observaciones dan como resultado los mismos juicios. Un indicador que se utiliza comúnmente para la confiabilidad entre jueces es la *kappa de Cohén*.

Confiabilidad y precisión de las mediciones

La confiabilidad es obviamente importante cuando los investigadores elaboran medidas del comportamiento. La confiabilidad no es la única característica de una medida ni el único aspecto por el que se preocupan los investigadores; nos habla del error de medición, pero no nos indica si tenemos una medida adecuada de la variable de interés. Para emplear un ejemplo tonto, suponga que deseamos medir la inteligencia. La medida que elaboramos se parece mucho al aparato que se emplea para medir la talla de los zapatos en la zapatería local. Entonces le pedimos a usted que coloque su pie en el aparato y utilizamos el calibrador para medir su inteligencia. Una escala de inteligencia proporciona números, de modo que se evalúa de inmediato el grado de inteligencia de una persona. ¿Darán estos números el resultado de una medida confiable de la inteligencia? ¡La respuesta es que sí lo harán! Consideren lo que sucedería con el coeficiente de confiabilidad de *test-retest.* Si aplico la "escala de inteligencia del pie" el lunes, será casi igual el siguiente lunes; la confiabilidad de *test-retest* es alta. Pero, ¿se trata de una medida precisa de la inteligencia? Obviamente, las puntuaciones no tienen ninguna relación con la inteligencia; el simple hecho de que hayamos nombrado al aparato como prueba de inteligencia no significa que sea una medida correcta de ésta.

Consideremos un ejemplo menos tonto. Suponga que la bomba de la estación de gasolina de su vecindario pone la misma cantidad de gasolina en su auto-

móvil cada vez que compra un galón (o un litro) de combustible; la bomba de gasolina es confiable. Sin embargo, el tema de la precisión continúa abierto. La única forma en que usted puede conocer la precisión de la bomba es comparando el galón (o litro) que recibe con alguna medida estándar de un galón. De hecho, las autoridades respectivas tienen inspectores que son responsables de comparar la cantidad que la bomba indica como un galón con la unidad de medida exacta que le corresponde. La bomba que no despacha lo que señala, debe repararse o reemplazarse. Esta diferencia entre la confiabilidad y la precisión de las mediciones nos lleva a estudiar la validez de las mediciones.

VALIDEZ DE CONSTRUCTO DE LAS MEDICIONES

Si algo es válido, es "verdadero" en el sentido de que está apoyado por evidencias disponibles. La cantidad de gasolina que indica la bomba debe coincidir con alguna medida estándar del volumen de líquidos; la medida de una característica de la personalidad, como la timidez, debe ser un indicador preciso de dicho rasgo. En el capítulo 4 indicamos que la **validez de constructo** se refiere a qué tan adecuadas son las definiciones operacionales de las variables. ¿Qué tanto refleja realmente la definición operacional de una variable su significado teórico verdadero? En términos de medición, la validez de constructo se refiere a si la medida que se emplea en realidad mide el constructo que se supone debe medir. A algunos individuos que solicitan empleo se les pide que respondan la prueba de habilidades administrativas; se supone que esta medida predice las habilidades administrativas de una persona. La validez de una prueba como ésta se determina si en realidad mide esta habilidad. Una medida de la timidez es una definición operacional de la variable timidez; la validez de esta medida se determina si mide este constructo.

Indicadores de la validez de constructo

¿Cómo sabemos si una medida es válida? En la figura 5.2 se resumen las formas en que podemos evaluar la validez. La forma más sencilla para afirmar que una medida es válida es sugerir que la medida evalúa, aparentemente con precisión, la variable en cuestión. A esto se le llama **validez aparente** —la evidencia de validez es que la medida "parece" medir lo que se supone debe medir—. La validez aparente no es muy sofisticada; implica únicamente el juicio de que, dada la definición teórica de la variable, el contenido de la medida aparentemente mide la variable. Es decir, ¿los procedimientos utilizados para medir la variable parecen ser una definición operacional precisa de la variable teórica? Así, una medida de una variable, como la timidez, generalmente parecerá medir dicha variable. Una medida de la timidez, denominada la Q tímida (Bortnik, Henderson y Zimbardo, 2002), incluye reactivos como "con frecuencia me siento inseguro en situaciones sociales", pero no incluye reactivos como "aprendí a andar en bicicle-

Validez aparente

El contenido de la medida aparentemente refleja el constructo que se está midiendo.

Validez orientada al criterio

Las puntuaciones de la medida están relacionadas con un criterio (un indicador del constructo).

Tipos de validez orientada al criterio

Validez predictiva	Las puntuaciones de la medida predicen el comportamiento con base en un criterio.
Validez concurrente	Los individuos que pertenecen a grupos que ya se sabe difieren en el constructo, obtienen distintos puntajes en la medida.
Validez convergente	Los puntajes de la medida están relacionadas con otras medidas del mismo constructo.
Validez discriminante	Los puntajes de la medida *no* están relacionados con otras medidas que son teóricamente diferentes.

FIGURA 5.2
Indicadores de la validez de constructo de una medida.

ta a temprana edad", el primer tipo parece estar más relacionado con la timidez que el segundo.

No obstante, la validez aparente no es suficiente para concluir que una medida sea realmente válida. La apariencia no es un indicador muy bueno de la precisión. Algunas medidas pobres pueden tener validez aparente; por ejemplo, la mayoría de las medidas de la personalidad que aparecen en revistas populares suelen incluir varias preguntas que parecen razonables, pero generalmente no le dicen algo significativo. La interpretación de los puntajes puede ser divertida, mas no existe evidencia empírica que sustente las conclusiones que se plantean en el artículo. Además, muchas medidas correctas de variables no poseen una validez aparente que sea obvia. Por ejemplo, ¿será obvio que los movimientos oculares rápidos cuando está durmiendo son una medida de la aparición de los sueños? En lugar de confiar en la validez aparente, los investigadores evalúan la validez al realizar investigaciones con la medida.

La investigación sobre la validez de constructo examina la relación entre los puntajes de una medida y algún criterio; a esto Trochim (2000) lo llamó **validez orientada al criterio.** Un investigador hace una predicción, con base en la teo-

ría que subyace al constructo, de la forma en que el desempeño en la medida se relacionará con otra variable llamada criterio. Existen cuatro tipos de métodos de investigación, relacionados con el criterio, que difieren según el tipo de criterio que se emplea: la validez predictiva, la validez concurrente, la validez convergente y la validez discriminante (Trochim, 2000). Estas se relacionan entre sí debido a que todas constituyen formas para demostrar la validez de constructo de una medida.

Validez predictiva

La **validez predictiva** es el grado en que la medida permite pronosticar las conductas que debe predecir. Por lo tanto, en la validez predictiva el criterio es alguna conducta futura. En el caso de algunas medidas, la naturaleza del criterio es obvia. Una prueba de habilidades administrativas tiene el objetivo de proporcionar una forma rápida para predecir el desempeño futuro en un puesto administrativo. De manera similar, muchos estudiantes universitarios responden el *Gradúate Record Exam* (GRE), que se elaboró para predecir el éxito en programas de posgrado, o el *Law School Admission Test* (LSAT), creado para predecir el éxito en la Escuela de leyes. La validez de constructo de medidas como éstas se demuestra cuando los puntajes de la medida sí predicen el comportamiento futuro. Por ejemplo, la validez predictiva del LSAT se demuestra cuando la investigación indica que las personas que obtienen puntajes altos en la prueba tienen mayor éxito en la escuela de leyes que las personas que obtienen puntajes bajos (por ejemplo, existe una relación positiva entre el puntaje en la prueba y las calificaciones en la escuela de leyes). La medida puede utilizarse para aconsejar a los individuos con respecto al éxito que tendrán en la escuela de leyes o para seleccionar a quienes solicitan entrar a esta escuela.

Para otros tipos de variables, las predicciones son menos obvias. Si tiene una medida de la timidez, puede hacer predicciones sobre la manera en que los puntajes de la medida se relacionarán con diversas conductas. Por ejemplo, su teoría de la timidez podría pronosticar que los puntajes de timidez se relacionarán con la cantidad de ansiedad que las personas mostrarán cuando se describan a sí mismas ante un extraño. Si esta predicción resulta correcta, aumenta su confianza en la validez de la medida. Sin embargo, la confirmación de una sola predicción nunca se utilizaría como evidencia única para la validez de constructo de una medida de una variable, como la timidez. Mejor dicho, la validez de la medida se construye con el tiempo, conforme diversos estudios investigan la teoría del constructo que se está midiendo. Además, las medidas de las variables suelen tener una vida limitada. Conforme se acumulan los hallazgos de investigaciones, los científicos refinan la medida y en ocasiones crean medidas nuevas que hacen mejores predicciones. Este proceso conduce a mejores medidas y a una comprensión más completa de la variable subyacente en estudio.

Validez concurrente

El criterio para la investigación de la **validez concurrente** es el hecho de que dos o más grupos de personas difieran, de la manera esperada, en la medida. Regresando al ejemplo de la medida de la timidez, su teoría de la timidez podría hacer que usted esperara que vendedores cuyo trabajo requiere hacer "llamadas frías" a clientes potenciales, obtuvieran menores puntajes en su escala de timidez que los vendedores que trabajan en puestos que requieren tratar únicamente con personas que han hecho el esfuerzo para contactar a la empresa, con el fin de hacer una compra potencial.

Validez convergente

Cualquier medida es una definición operacional particular de la variable en estudio. Con frecuencia existen otras definiciones operacionales —otras medidas— del mismo constructo o de constructos similares. La **validez convergente** es el grado en que los puntajes de la medida en cuestión se relacionan con puntajes de otras medidas del mismo constructo o de constructos similares. Las medidas de constructos similares deben "converger"; por ejemplo, una medida de la timidez debe tener una alta correlación con otra medida de la timidez o una medida de un constructo similar, como la ansiedad social. En una investigación real, llevada a cabo con una escala de timidez, se demostró la validez de convergencia de la Q de timidez por medio de una correlación alta (0.77) entre los puntajes de esta escala y las de otra, denominada Temor de una evaluación negativa (Bortmk *et al,* 2002).

Validez discriminante

La **validez discriminante** se demuestra cuando la medida *no* está relacionada con variables con las que no debe estarlo. La medida debe discriminar entre el constructo que se mide y otros constructos que no tienen relación con él. Al investigar la validez discriminante de su medida de la timidez, Bortnik y sus colaboradores (2002) no encontraron relación entre los puntajes de la Q de timidez y varios valores interpersonales que no tienen una relación conceptual, como la evaluación de la fortaleza de los demás.

La Escala de búsqueda de sensaciones (Zuckerman, 1979) es un excelente ejemplo de la investigación de la validez de constructo, en la que diversos estudios sustentan las predicciones hechas por una teoría. La investigación de Zuckerman se *vio* motivada por una teoría psicológica sobre los niveles óptimos de activación fisiológica. La teoría establece que las personas necesitan mantener un nivel óptimo de activación. Cuando la activación es demasiado baja, los individuos se sienten motivados a realizar cosas para incrementar la activación;

cuando ésta es demasiado alta, los individuos tienden a reducirla. La teoría ayuda a explicar muchas conductas, como las alucinaciones y otras perturbaciones que la gente experimenta cuando se le coloca en ambientes con privación sensorial. Zuckerman decidió estudiar esta teoría enfocándose en las diferencias individuales; se preguntó por qué muchas personas parecen buscar de manera consistente sensaciones novedosas o activadoras (por ejemplo, el paracaidismo, escuchar música a un volumen alto, conducir automóviles de carreras), mientras que otros individuos evitan las sensaciones de activación.

La Escala de búsqueda de sensaciones se creó para estudiar este tipo de diferencias individuales en la personalidad. La escala incluye reactivos que tratan de medir la búsqueda de emociones, la susceptibilidad al aburrimiento y otros aspectos de la búsqueda de sensaciones. Por supuesto, se midió la confiabilidad de la escala; después de determinar que era confiable, se pudo iniciar la investigación sobre su validez de constructo. Durante un periodo de muchos años, las investigaciones de Zuckerman y otros autores demostraron la validez predictiva: las personas que obtienen puntajes altos en la escala se comportan en forma diferente que quienes obtienen puntajes bajos. Por ejemplo, los individuos que buscan sensaciones realizan actividades más peligrosas, conducen a mayor velocidad y prefieren actividades menos intelectuales. Otro estudio en el que se utilizó la Escala de búsqueda de sensaciones, examinó la relación de esta medida con otros constructos similares y diferentes de la personalidad. El trabajo con esta escala ha llevado a la investigación de las bases biológicas de la búsqueda de sensaciones, como la identificación de los mecanismos cerebrales que son responsables de las necesidades de activación y las bases genéticas de este rasgo La investigación de Zuckerman ilustra un programa sistemático de investigación sobre la validez de la medida de un constructo psicológico.

REACTIVIDAD DE LAS MEDICIONES

Un problema potencial que surge cuando se mide el comportamiento es la **reactividad.** Se dice que una medida es reactiva, si el hecho de que un individuo sea consciente de que está siendo medido cambie su conducta. Una medida reactiva señala cómo es la persona cuando se da cuenta de que está siendo observada, pero no indica la forma en que se comportaría en circunstancias naturales El simple hecho de tener diversos aparatos conectados al cuerpo, como electrodos o un instrumento para medir la presión sanguínea, puede cambiar las respuestas fisiológicas que se registran. Saber que un investigador está observando o registrando su conducta en una cinta, puede modificar la forma en que se comporta. Las medidas del comportamiento varían en términos de su reactividad potencial. También existen formas para minimizar la reactividad, come dar tiempo a los individuos para que se acostumbren a la presencia del observador o del equipo de registro.

Un libro escrito por Webb, Campbell, Schwartz, Sechrest y Grove (1981) ha centrado la atención en diversas medidas denominadas *no reactivas* o *no direc-*

tivas. Muchas de estas medidas incluyen formas inteligentes del registro indirecto de una variable. Por ejemplo, una medida no directiva sobre las preferencias de las pinturas en un museo de arte es la frecuencia con que deben cambiarse las baldosas que rodean a cada pintura; las pinturas más populares son aquellas con cambios más frecuentes. El estudio de Levine (1990) sobre el ritmo de vida en las ciudades, mencionado en el capítulo 2, empleó medidas indirectas como la precisión de los relojes en los bancos y la velocidad del procesamiento de las solicitudes estándares en las oficinas de correos, para medir el ritmo de vida. Algunas de las medidas descritas por Webb y sus colaboradores (1981) son cómicas. Por ejemplo, en 1872, Sir Francis Galton estudió la *eficacia de* rezar para tener una vida larga. Galton se cuestionó si la realeza británica, que con frecuencia era el recipiente de los rezos del pueblo, vivía más que otras personas. Verificó registros de las muertes y encontró que los miembros de las familias reales en realidad vivían menos que los demás, al igual que los literatos y los científicos. El libro de Webb y sus colaboradores es una rica fuente de medidas no reactivas como éstas. Además, logra dirigir la atención hacia el problema de la reactividad y sensibiliza a los investigadores de la necesidad de reducir la reactividad, siempre que sea posible. Regresaremos a este tema en diversos puntos de este texto.

VARIABLES Y ESCALAS DE MEDICIÓN

Cada variable que se estudie debe definirse operacionalmente. La definición operacional es el método específico utilizado para manipularla o medirla (véase el capítulo 4). Debe haber, al menos, dos valores o niveles de la variable. En el capítulo 4 mencionamos que los valores pueden ser cuantitativamente diferentes o pueden reflejar diferencias categóricas. En realidad, el mundo es un poco más complejo. Los niveles pueden conceptualizarse en una escala que utilice uno de cuatro tipos de escalas de medición: nominal, ordinal, de intervalo y de razón.

Escalas nominales

Las **escalas nominales** no poseen propiedades numéricas o cuantitativas. En su lugar, las categorías o grupos sencillamente difieren entre sí (en ocasiones las variables nominales se conocen como variables "categóricas"). Un ejemplo obvio es la variable del género: a una persona se le clasifica como hombre o mujer. Ser hombre no implica una mayor cantidad de "género" que ser mujer; los dos niveles son simplemente distintos. Se le llama escala nominal debido a que sólo asignamos nombres a las distintas categorías. Otro ejemplo es la clasificación de los estudiantes universitarios de acuerdo con la carrera que estudian. Por ejemplo, no podría dársele un número mayor a la carrera de psicología que a la de historia, incluso si se asignan números a las distintas categorías, los números carecerían de significado, excepto como una forma de identificación.

En un experimento, con frecuencia la variable independiente es de tipo nominal o categórico. Por ejemplo, Punnett (1986) estudió una variable que po-

dría denominarse "tipo de motivación". Se alentó a trabajadores con una meta de desempeño específica y difícil o un fin vago expresado como "haga lo mejor que pueda". La variable del establecimiento de la meta es claramente de tipo nominal, ya que los dos niveles son sólo distintos; las metas no tienen propiedades numéricas. Punnett encontró que la meta específica producía un mejor desempeño que el objetivo vago.

Escalas ordinales

Las **escalas ordinales** nos permiten ordenar los niveles de la variable en estudio. En lugar de tener categorías que sólo son diferentes, como en una escala nominal, las categorías pueden ordenarse de la primera a la última. Un ejemplo de esta escala es el sistema de clasificación de filmes que se utiliza en la sección de televisión y espectáculos del periódico local. Las películas que pasan en televisión reciben una, dos, tres o cuatro estrellas, con base en las siguientes descripciones:

* * * *	Nueva o vieja, clásica
* * *	De primera
* *	Mala; puede tener buenos momentos
*	Para desesperarse

El sistema de orden no es una escala nominal debido a que el número de estrellas es significativo, en términos de un continuo de la calidad. Sin embargo, las estrellas únicamente nos permiten ordenar las películas. Un filme con cuatro estrellas es mejor que uno con tres; una película con tres estrellas es mejor que una con dos, y así sucesivamente. Aun cuando tengamos esta información cuantitativa sobre las películas, no podemos decir que la diferencia entre una película con una estrella y otra con dos estrellas sea siempre la misma o que sea igual a la diferencia entre una película con dos estrellas y otra con tres estrellas. No hay un valor particular vinculado con los intervalos entre los números empleados en la escala ordenada.

Escalas de intervalo y de razón

En una **escala de intervalo,** la diferencia entre los números tiene un significado. Específicamente, los intervalos entre los números son del mismo tamaño. Por ejemplo, la diferencia entre uno y dos en la escala es la misma que entre dos y tres. Las escalas de intervalo generalmente tienen *cinco* o más niveles cuantitativos. Un termómetro casero (Fahrenheit o Celsius) mide la temperatura en una escala de intervalo. La diferencia de temperatura entre 40° y 50° es igual a la diferencia que existe entre 70° y 80°. No obstante, no existe un cero absoluto en la escala que indique la ausencia de temperatura. El cero en cualquier escala de intervalo es únicamente un punto de referencia arbitrario. Con la f alta de un

cero absoluto, no podemos obtener cocientes de los números. Es decir, no podemos afirmar que un número en la escala representa el doble (o el triple, etc.) de Temperatura que otro número. Por ejemplo, usted no puede decir que 60° implica el doble de calor que 30°.

Un ejemplo de una escala de intervalo en las ciencias del comportamiento sería la medida de un rasgo de la personalidad, como la extroversión. Si la medida está en una escala de intervalo, no podemos afirmar cosas como "la persona que obtuvo una puntuación de 20 es dos veces más extrovertida que la persona que obtuvo una puntuación de diez", ya que no existe un cero absoluto que indique la ausencia del rasgo medido.

Las **escalas de razón** poseen un cero absoluto que indica la ausencia de la variable que se mide. Algunos ejemplos incluyen muchas medidas físicas, como la longitud, el peso o el tiempo. Con una escala de razón es posible hacer afirmaciones tales como que "una persona que pesa 100 kilos tiene el doble de peso que otra que pesa 50", o que "los participantes del grupo experimental respondieron dos veces más rápido que los del grupo control".

Las escalas de razón se utilizan en las ciencias del comportamiento cuando se estudian variables que implican medidas físicas, en particular medidas de tiempo tales como el tiempo de reacción, la tasa de respuesta y la duración de la respuesta. Sin embargo, muchas variables en las ciencias del comportamiento son menos precisas, por lo que emplean escalas de medición nominales, ordinales o de intervalo. También debe señalarse que las pruebas estadísticas para las escalas de intervalo y de razón son las mismas.

La importancia de las escalas de medición

Cuando usted lea *acerca* de las definiciones operacionales de variables, reconocerá los niveles de las variables en términos de estos tipos de escalas. Las conclusiones que se hacen sobre el significado de un puntaje específico de una variable dependen del tipo de escala que se utilice. Con las escalas de intervalo y de razón se pueden hacer distinciones cuantitativas que permiten hablar sobre cantidades de la variable. Con las escalas nominales no existe información cuantitativa. Por ejemplo, suponga que está estudiando la percepción del atractivo físico. En un experimento podría mostrar a los participantes fotografías de personas con diferentes características, tales como el cociente de su cintura y su cadera (tamaño de la cintura dividido entre el tamaño de la cadera). Singh (1993) ha estudiado extensamente esta variable. ¿Cómo mediría los juicios que hacen los participantes del atractivo físico? Podría utilizar una escala nominal tal como:

_____ No atractivo _____ Atractivo

Estos valores de la escala permiten que los participantes digan si consideran que la persona es o no atractiva, pero no permiten conocer la cantidad del atractivo. Como alternativa, se podría utilizar una escala que pida a los participantes juzgar la cantidad del atractivo:

Nada atractivo ___ ___ ___ ___ ___ ___ ___ Muy atractivo

Esta escala proporciona la información cuantitativa sobre la cantidad del atractivo, ya que se pueden asignar valores numéricos a cada una de las opciones de respuesta en la escala; en este caso, los valores irían del uno al siete. Uno de los principales hallazgos de la investigación de Singh es que los hombres consideran que las mujeres con un cociente de 0.70 de la cintura y la cadera son las más atractivas. Singh interpreta estos resultados en términos de la teoría de la evolución; quizás este cociente sea una señal de la capacidad reproductiva.

La escala que se utiliza también determina el tipo de estadística que es adecuada para el análisis de los resultados de un estudio. Por ahora, no necesitamos preocuparnos por los análisis estadísticos. No obstante, regresaremos a este tema en el capítulo 12.

Ahora estamos preparados para estudiar métodos para la medición del comportamiento. En el capítulo 6 se describe una variedad de métodos de observación. Después, en el capítulo 7, nos enfocaremos en los cuestionarios y las entrevistas.

Términos de estudio

Alfa de Cronbach
Coeficiente de correlación producto-momento de Pearson Confiabilidad
Confiabilidad de consistencia interna
Confiabilidad de *test-retest*
Confiabilidad entre jueces Confiabilidad por mitades Correlación reactivo-total
Error de medición Escala de intervalo
Escala de razón
Escala nominal Escala ordinal
Puntuación verdadera
Reactividad Validez aparente
Validez concurrente Validez convergente Validez de constructo Validez discriminante Validez orientada al criterio Validez predictiva Variable criterio

Preguntas de repaso

1. *¿A* qué se refiere la confiabilidad de una medida? Distinga entre el puntaje verdadero v el error de medición.
2. Describa los métodos para determinar la confiabilidad de una medida.
3. Comente el concepto de validez de constructo. Establezca la diferencia entre la validez convergente y la discriminante.
4. *¿A* qué se refiere la validez orientada al criterio?
5. ¿Por qué no es suficiente la validez aparente para establecer la validez de una medida?

6. ¿Qué es una medida reactiva?
7. Establezca la diferencia entre las escalas nominales, ordinales, de intervalo y de razón.

Actividades

1. Encuentre un libro sobre la medición en psicología, como el de Robinson, Shaver y Wrightsman (1991) o realice una búsqueda bibliográfica sobre la validez de constructo. Identifique una medida que le interese y describa la confiabilidad y la validez reportadas para esta medida.
2. A continuación se incluyen algunas variables. Para cada una, identifique si se utiliza una escala nominal, ordinal, de intervalo de razón:
   a) Las temperaturas en las ciudades del país que se incluyen en la mayoría de los periódicos.
   b) Los pesos de bebés recién paridos en el Wilshire General Hospital la semana pasada.
   c) La cantidad de horas que pasó estudiando cada día la semana pasada.
   d) El monto de la propina que se deja después de una comida de tres horas de duración en un restaurante.
   e) La cantidad de votos recibidos por los candidatos republicanos y demócratas al Congreso, en su distrito electoral en la elección del año 2000.
   f) La marca de aparatos para *DVD* que ocupa el tercer lugar en la evaluación de consumidores en una revista.
   g) El equipo de Connecticut recibió votos como el número uno en la encuesta de periodistas de deportes, y Kansas fue el número dos.
   h) La calificación de un amigo suyo en una prueba de inteligencia es de 109.
   i) El color de la pared de mi oficina es amarillo y el de la oficina del jefe es blanco.
   j) El tipo de programación en cada estación de radio de su ciudad (por ejemplo, la KPSY toca jazz, la KSOC es radio hablada).
3. Responda a una prueba de personalidad en Internet (pueden encontrar este tipo de pruebas por medio de los navegadores de Internet). Con base en la información que le proporcionaron, ¿qué puede concluir sobre la confiabilidad, la validez de constructo *y* la reactividad?
4. Piense en una característica importante que buscaría en un compañero romántico potencial, como simpático, inteligente, atractivo, trabajador, religioso, etc. ¿Cómo mediría esa característica? Describa dos métodos que podría utilizar para evaluar la validez de constructo.

6

Observación del comportamiento

Métodos cuantitativo y cualitativo

Observación naturalista
Descripción e interpretación de los datos
Aspectos de la observación naturalista
Participación y encubrimiento Definición de un ámbito de observación Límites de la observación naturalista

Observación sistemática
Sistemas de codificación
Aspectos metodológicos
Equipo
Reactividad
Confiabilidad
Muestreo

Estudios de caso

Investigación documental
Registros estadísticos
Archivos de encuestas
Registros escritos *y* de comunicación masiva
Análisis de contenido de documentos

Descripción de la personalidad y diferencias individuales

Términos de estudio

Preguntas de repaso

Actividades

El punto más importante de la investigación no experimental es la observación o medición del comportamiento. Dado que el comportamiento es tan variado y sucede en numerosos ambientes, los científicos sociales han desarrollado una cantidad considerable de formas para llevar a cabo investigación no experimental. Exploraremos una diversidad de métodos, entre ellos la observación del comportamiento en ambientes naturales, pedir a las personas que describan sus conductas (autorreporte) y el examen de registros de conducía ya existentes, como los datos de censos o archivos hospitalarios. Debido a que gran parte de la investigación no experimental incluye encuestas que utilizan cuestionarios o entrevistas, abordamos el tema de la investigación de encuesta en un capítulo separado. Antes de describir estos métodos con detalle, será útil comprender la diferencia entre los métodos cuantitativo y cualitativo de la descripción del comportamiento.

MÉTODOS CUANTITATIVO Y CUALITATIVO

Los métodos de observación pueden clasificarse, en general, como cuantitativos y cualitativos. Para comprender la diferencia, imagine que está interesado en describir la forma en que el trabajo afecta la vida de los adolescentes. Usted podría utilizar un método cuantitativo si elabora un cuestionario y le pide a un grupo de adolescentes que lo responda. Podría incluir preguntas sobre el número de horas que trabajan, el tipo de labor que realizan, sus estados de estrés, sus calificaciones escolares y el consumo de drogas. Después de asignar valores numéricos a las respuestas, puede someter los datos a un análisis estadístico o cuantitativo. La descripción cuantitativa de los resultados se enfocaría en as-rectos como el porcentaje de adolescentes que trabajan y la forma en que esta proporción varía de acuerdo con la edad. Algunos de los resultados de este tipo de encuestas se describen el capítulo 7.

Suponga, por otro lado, que utiliza un enfoque cuantitativo para describir el comportamiento. Podría organizar una serie de grupos de enfoque, con ocho o diez adolescentes en cada uno, y motivarlos a discutir sobre su percepción y experiencia con el mundo laboral. Podría pedirles que hablaran del tema con sus propias palabras y con base en sus marcos de referencia cognoscitivos. Para registrar las discusiones de los grupos de enfoque, podría utilizar un sistema audiovisual y hacer una transcripción posteriormente, o podría pedir a algunos observadores que tomaran notas detalladas durante la discusión. Una descripción cualitativa de los hallazgos se centraría en los temas que surgen en la polémica y en la manera en que los adolescentes conceptualizan estos temas. Este tipo de descripción es cualitativo, ya que está expresado en términos no numéricos, pero sí con el uso del lenguaje y las imágenes.

También se podrían emplear otros métodos, tanto cualitativos como cuanticativos para estudiar la actividad laboral de los adolescentes. Tenga en mente la diferencia entre los métodos cuantitativo y cualitativo para describir el comportamiento conforme lea sobre otros métodos de observación específicos que

se estudian en este capítulo. Ambos enfoques son valiosos y nos brindan distintas formas de comprensión.

OBSERVACIÓN NATURALISTA

En ocasiones, la observación naturalista se *conoce* como *trabajo de campo o* sencillamente *observación de campo* (véase Loñand y Lofland, 1995). En un estudio de **observación naturalista,** el investigador realiza observaciones en un ambiente natural en particular (el campo), durante mucho tiempo y por medio de diversas técnicas para reunir información. El reporte incluye estas observaciones y las interpretaciones del investigador con respecto de los hallazgos. Este método de investigación tiene su origen en la antropología y en el estudio del comportamiento animal; en la actualidad, se utiliza con frecuencia en las ciencias sociales para estudiar gran cantidad de fenómenos en todo tipo de ambientes sociales y organizativos.

El trabajo de Sylvia Scribner (1997) sobre el "pensamiento práctico" es un buen ejemplo de investigación de observación naturalista en la psicología. Scribner estudió las maneras en que las personas de diversas ocupaciones toman decisiones y resuelven problemas. Ella describe el proceso de esta investigación: "... mis colegas y yo estuvimos alrededor de una línea de producción de leche a las tres de la mañana, ayudamos a los cajeros a sumar los recibos y observamos a los operadores de la máquina realizar la producción del día. . . hicimos registros detallados de la forma en que las personas ejecutaban su trabajo. Reunimos copias de todo el material escrito que leen o producen; todo, desde las notas garabateadas en papel de estraza, hasta impresiones de computadora. Sacamos fotografías de los aparatos en su ambiente de trabajo que requerían del procesamiento de otros tipos de información simbólica, como termómetros, calibradores, escalas, instrumentos de medición de género diverso" (Scribner, 1997, p. 223). Un aspecto del pensamiento que estudió esta autora fue la manera en que los trabajadores realizan cálculos matemáticos. Encontró que los repartidores de leche, y otros trabajadores, realizan cálculos complejos que dependen de sus conocimientos adquiridos en ese trabajo. Por ejemplo, una factura de entrega podría implicar que el conductor multiplicara 32 cuartos de leche por 0.68. Para llegar a la respuesta, los conductores utilizan los conocimientos que han adquirido en el empleo sobre la cantidad de cuartos que caben en una caja y el costo de cada una de éstas; así, multiplican 2 cajas de leche por 10.88. En general, los trabajadores que Scribner observó emplearon estrategias complejas, aunque muy eficientes, para resolver los problemas en el trabajo. Algo muy importante es que las estrategias utilizadas con frecuencia no podían predecirse a partir de modelos formales de solución de problemas.

Un investigador utiliza la observación naturalista cuando desea describir y comprender la forma en que la gente vive, trabaja y experimenta un ambiente, ya sea social o cultural. Por ejemplo, si desea estudiar los bares como un ambiente social, necesita visitar uno o más bares durante mucho tiempo, hablar con la

gente, observar las interacciones y ser aceptado como un cliente "regular" (por ejemplo, Cavan, 1966). Si usted desea conocer cómo la gente persuade a otros, puede conseguir un trabajo como vendedor de coches o tomar un curso de entrenamiento para vendedores de enciclopedias (por ejemplo, Cialdini, 1988). Si usted está interesado en la forma en que las personas se convierten en parte de algún grupo social (por ejemplo, consumidores de mariguana, prostitutas o miembros de un culto religioso en particular), puede buscar la forma de reunirse con sus integrantes y entrevistarlos acerca de sus experiencias (por ejemplo, Becker, 1963, sobre consumidores de mariguana). Los investigadores que han estudiado lo que realmente es ser un paciente de un hospital mental, han asumido el papel de pacientes (por ejemplo, Rosenhan, 1973). Por supuesto, tal vez usted no desee hacer ninguna de estas cosas, sin embargo, si le interesan dichas cuestiones, los reportes escritos de estos investigadores constituyen una lectura fascinante.

Descripción e interpretación de los datos

La observación naturalista demanda que los investigadores se sumerjan en la situación. El investigador de campo observa todo: el ambiente, los patrones de las relaciones personales, la reacción de la gente ante los acontecimientos, entre otros elementos. El objetivo es ofrecer una imagen completa y precisa, más que probar hipótesis formuladas antes del estudio. Para lograr esta meta, el investigador debe tomar notas de campo detalladas; es decir, escribir o dictar con regularidad (al menos una vez cada día) todo lo que ha sucedido. Los investigadores de campo utilizan una variedad de técnicas para reunir información: observan a las personas y los acontecimientos, entrevistan a "informantes" clave que proporcionan información, hablan con las personas sobre su vida y examinan documentos producidos en el ambiente, como periódicos, cartas o memorándums. Además de tomar notas de campo detalladas, los investigadores que realizan observaciones naturalistas generalmente emplean registros audiovisuales.

La primera meta del investigador es describir el ambiente, los sucesos y las personas observadas. La segunda, igualmente importante, es analizar lo observado. El investigador debe interpretar lo que ocurrió, esencialmente generando hipótesis que lo ayuden a explicar los datos y a hacerlos comprensibles. Este tipo de análisis se realiza construyendo una estructura coherente para describir las observaciones. El reporte final, aunque sensible al orden cronológico de los acontecimientos, suele estar organizado alrededor de la estructura desarrollada por el investigador. Los ejemplos específicos de hechos que suceden durante la observación se utilizan para sustentar las interpretaciones del científico.

Un reporte satisfactorio de observación naturalista sustenta el análisis por medio del uso de múltiples confirmaciones. Por ejemplo, sucesos similares pueden ocurrir varias veces, dos o más personas pueden reportar información similar y pueden ocurrir muchos acontecimientos diferentes que sustenten, todos, la misma conclusión.

Los datos de las observaciones naturalistas son principalmente de carácter *cualitativo;* es decir, son las descripciones de las observaciones mismas y no los

resúmenes estadísticos *cuantitativos.* Este tipo de descripciones cualitativas con frecuencia es más rico y se acercan más al fenómeno en estudio que las representaciones estadísticas. Sin embargo, no hay razón por la que no se puedan recolectar datos cuantitativos en un estudio de observación naturalista. Si las circunstancias lo permiten, se pueden obtener datos con respecto a los ingresos, el tamaño de la familia, el nivel educativo y otras variables fáciles de cuantificar. Estos datos pueden reportarse e interpretarse junto con los datos cualitativos recolectados por medio de entrevistas y observaciones directas.

Aspectos de la observación naturalista

Participación y encubrimiento Dos aspectos relacionados que debe enfrentar el investigador son: si debe o no ser un participante del ambiente social, y si necesita ocultar sus propósitos de las demás personas en el ambiente. ¿Se convertirá en un participante activo del grupo u observará desde el exterior? ¿Ocultará sus propósitos e incluso su presencia o permitirá abiertamente que las personas sepan lo que está haciendo?

Un observador que no participa es un personaje externo que no forma parte activa del ambiente. En contraste, un observador participante toma un papel interno activo. Debido a que la observación participante permite al investigador observar el ambiente desde adentro, éste puede ser capaz de experimentar los acontecimientos de la misma forma que los participantes naturales. Las amistades y otras experiencias del observador participante pueden brindar datos valiosos. No obstante, un problema potencial de la observación participante es que el observador puede perder la objetividad necesaria para realizar observaciones científicas. El hecho de continuar siendo objetivo puede ser especialmente difícil cuando el investigador ya pertenece al grupo de estudio (por ejemplo, un investigador que pertenece al grupo de Padres sin pareja y lleva a cabo un estudio en ese grupo). Recuerde que la observación naturalista requiere descripciones precisas e interpretaciones objetivas, sin hipótesis previas. Si un investigador tiene razones previas para criticar a las personas en el ambiente c para hacer un reporte elogioso sobre un grupo en particular, es muy probable que las observaciones estén sesgadas y que las conclusiones carezcan de objetividad.

¿Debe el investigador ocultar o comunicar los propósitos del estudio? A veces es preferible la observación oculta, ya que la presencia del observador puede influir y alterar el comportamiento de los individuos observados. Imagine cómo la presencia de un observador visible podría alterar el comportamiento de estudiantes de preparatoria en muchas situaciones escolares. Por lo tanto, 1a observación encubierta es menos reactiva que la abierta, porque las personas no saben que se está observando y registrando su comportamiento. Sin embargo, k observación abierta puede ser preferible desde un punto de vista ético: ¡considere la invasión de la intimidad si los investigadores se esconden debajo de las camas en los dormitorios, para descubrir de qué hablan los estudiantes universitarios! (Henle y Hubbell, 1938). Además, las personas suelen acostumbrarse

rápidamente al observador y a comportarse de manera natural ante su presencia. Dos ejemplos muy conocidos de observación abierta se dieron en la televisión. En la serie documental de la PBS *Una historia de amor estadounidense,* realizada en 1999, y en el programa *Mundo real,* de MTV, se filmó a las personas viviendo juntas durante un largo periodo. Muchos espectadores de estos programas se sorprenden al ver cuan rápidamente las personas se olvidan de las cámaras y revelan de manera espontánea muchos aspectos privados de su vida.

La decisión de ocultar o no los propósitos o la presencia del investigador depende tanto de los aspectos éticos como de la naturaleza del grupo y el ambiente de estudio, en particular. En ocasiones, un observador participante no se oculta ante ciertos miembros de la congregación, quienes le dan permiso de formar parte de la comunidad como observador encubierto. También es común que un observador oculto tome la decisión de no revelar nada sobre sus propósitos inicialmente, pero si alguien se lo pide, le da toda la información acerca de las metas del estudio. Los observadores no participantes no se ocultan cuando obtienen permiso para "moverse" en un ambiente o para emplear técnicas de entrevistas con el fin de reunir información (por ejemplo, en el estudio de Becker sobre los consumidores de mariguana, algunas de las personas que fueron entrevistadas al principio introdujeron a Becker a su círculo de amigos que también consumían mariguana). Entonces, en realidad existen ambientes de participación y de encubrimiento: por ejemplo, es probable que un observador no participante nunca llegue a convertirse en un miembro del grupo, pero que con el tiempo sea aceptado como amigo o simplemente como parte de sus actividades. En suma, los investigadores que utilizan la observación naturalista para estudiar el comportamiento deben determinar cuidadosamente cuál será su papel en el ambiente.

Definición de un ámbito de observación Quizás el investigador que utiliza la observación naturalista quiera estudiar *todo* sobre un ambiente. Sin embargo, esto puede ser imposible, sencillamente debido a que un ambiente y las preguntas que pueden plantearse sobre éste son demasiado complejos. Por consiguiente, con frecuencia los investigadores deben limitar el ámbito de sus observaciones a conductas que sean relevantes para los temas centrales del estudio. Anteriormente mencionamos el interés de Cialdini en la influencia social en ámbitos como la venta de automóviles. En este caso, Cialdini podría enfocarse únicamente en las técnicas de venta e ignorar aspectos como las prácticas administrativas y las relaciones entre los vendedores.

Límites de la observación naturalista Es obvio que la observación naturalista no puede emplearse para estudiar todos los temas o fenómenos. Este método es más útil cuando se estudian ambientes sociales complejos, con el fin de comprenderlos y elaborar teorías basadas en las observaciones. Su utilidad es menor cuando se estudian hipótesis bien definidas, bajo condiciones especificadas con precisión. La investigación de campo también es muy difícil de realizar (véase Green y Wallaf, 1981). A diferencia de un experimento típico de laboratorio, la recolec-

ción de datos por medio de la investigación de campo no siempre puede programarse en un momento y lugar convenientes. De hecho, la investigación de campo puede ser extremadamente consumidora de tiempo y, con frecuencia, coloca al investigador en un ambiente poco familiar durante largos periodos. Además, en la investigación experimental los procedimientos están definidos correctamente y son los mismos para cada participante, y el análisis de datos se planea con anticipación. No obstante, en la observación naturalista existe un cambio constante en el patrón de acontecimientos, algunos importantes y otros sin importancia; el científico debe registrarlos todos y ser flexible para ajustarse a ellos conforme la investigación progresa. Finalmente, el proceso de análisis que se lleva a cabo con el término de la investigación no es sencillo. El investigador debe revisar una y otra vez los datos para crear hipótesis que los expliquen, y después asegurarse de que todos los datos sean consistentes con las hipótesis.

Si algunas de las observaciones no son consistentes, el especialista realiza más análisis, judd, Smith y Kidder (1991) enfatizan la importancia del **análisis de casos negativos.** Un caso negativo es una observación que no se ajusta a la estructura explicativa diseñada por el investigador. Cuando encuentra un caso negativo, el científico revisa las hipótesis y examina nuevamente todos los datos para asegurarse de que sean consistentes con la nueva hipótesis. Incluso, se pueden obtener datos adicionales para examinar más de cerca las circunstancias que condujeron al caso negativo. Aun cuando la investigación que emplea la observación naturalista implica un procedimiento difícil y desafiante, produce conocimientos invaluables cuando se realiza de forma adecuada.

OBSERVACIÓN SISTEMÁTICA

La **observación sistemática** se refiere a la observación cuidadosa de una o más conductas específicas en un ambiente en particular. Este método de investigación es mucho menos global que la observación naturalista. El investigador se interesa únicamente en algunos comportamientos específicos, las observaciones son cuantificables y es común que el científico haya elaborado hipótesis previas con respecto de las conductas.

Por ejemplo, Bakeman y Brownlee (1980; véase también Bakeman y Gottman. 1986) se interesaron en el comportamiento social de niños pequeños. Filmaron a niños de tres años de edad en una habitación, en una situación de "juego libre". A cada pequeño se le filmó durante 100 minutos; los observadores vieron las películas y codificaron la conducta de cada niño cada 15 segundos, utilizando el siguiente sistema de codificación:

Desocupado: el niño no está haciendo algo en particular o simplemente esta observando a otros niños.

Juego solitario: el niño juega solo, con juguetes, pero no está interesado ni se ve afectado por las actividades de los otros niños.

Reunido: el niño está con otros infantes, pero no está ocupado en alguna actividad en particular.

Juego paralelo: el niño juega junto a otros niños con juguetes similares, pero no está jugando con los otros.

Juego grupal: el niño juega con otros pequeños, comparte juguetes o participa en actividades de juego organizado como parte de un grupo de niños.

Bakeman y Brownlee estaban especialmente interesados en la secuencia u orden de las diferentes conductas en las que se involucraban los niños. Encontraron, por ejemplo, que en pocas ocasiones los menores pasaban de estar desocupados al juego paralelo. Sin embargo, con frecuencia iban del juego paralelo al grupal, lo cual indica que el juego paralelo es un estado de transición en el que los niños deciden continuar e interactuar en una situación grupal o *no.*

Sistemas de codificación

Muchas conductas pueden estudiarse por medio de la observación sistemática. El investigador debe decidir qué comportamientos le interesan, elegir un ambiente en el que se puedan observar las conductas y, muy importante, crear un **sistema de codificación,** como el que se describió anteriormente, para medir las conductas. En ocasiones el investigador elabora el sistema de codificación para que se ajuste a las necesidades del estudio en particular. Los sistemas de codificación deben ser lo más sencillos posibles, para que permitan, a los observadores, clasificar con facilidad las conductas. La necesidad de sencillez es especialmente importante cuando los observadores codifican conductas en *vivo,* más que cuando observan filmaciones que pueden revisarse o incluso codificarse cuadro por cuadro. Un ejemplo de un sistema de codificación sencillo proviene de un estudio realizado por Barton, Bal tes y Orzech (1980), en el cual se observó a los residentes y el personal de un asilo. Se utilizaron únicamente cinco categorías: 1) conducta independiente del residente (por ejemplo, hacer algo solo, como asearse); 2) conducta dependiente del residente (pedir ayuda); 3) conducta del personal que promueve la independencia (elogios o motivación de la independencia); 4) conducta del personal que promueve la dependencia (ayudar o motivar la aceptación de la ayuda), y 5) otras: conductas no relacionadas de los residentes o del personal. Sus resultados ejemplifican uno de los problemas de las instituciones que ofrecen cuidados personales: el personal se percibe como "cuidador", por lo que con frecuencia se conduce de forma tal que promueve las conductas de dependencia. ¿Guía esta conducta a una mayor dependencia de los residentes y, tal vez, a la pérdida de la sensación de control? Si es así, las con-secuencias pueden ser graves: recuerde el experimento de Rodin y Langer (1977), que se comentó en el capítulo 4, en el que la sensación de control condujo a una mayor felicidad y bienestar general entre los residentes de un asilo.

En ocasiones, los investigadores pueden utilizar sistemas de codificación elaborados por otros autores. Por ejemplo, el Sistema de codificación de interacción familiar (SCIF, véase Patterson y Moore, 1979) consiste en 29 categorías de interacción, las cuales están agrupadas como aversivas (hostilidad), prosociales (ayuda) y actividades generales. La mayor parte de la investigación que ha empleado el SCIF se ha centrado en la manera en que las conductas hostiles infantiles se aprenden y mantienen en una familia. Otra forma de codificación es el Sistema multinivel para la observación de grupos (SYMLOG, Bales y Cohén, 1979). El SYMLOG ofrece una forma para codificar la interacción entre individuos en grupos, sobre tres dimensiones principales: no amistoso-amistoso, expresivo emocionalmente-controlado instrumentalmente, y sumiso-dominante. Una de las principales ventajas de utilizar un sistema de codificación creado previamente, es que ya existe un cuerpo de investigación en el que el sistema ha demostrado ser útil, y que generalmente se dispone de materiales para el entrenamiento.

Aspectos metodológicos

Equipo Debemos mencionar brevemente varios aspectos metodológicos de la observación sistemática. El primero se refiere al equipo. Usted puede observar directamente la conducta y codificarla al mismo tiempo, por ejemplo, puede observar y registrar directamente el comportamiento de los niños en un salón de clases o de las parejas cuando interactúan en el *campus,* a través de medidas apuntadas en papel y lápiz. No obstante, cada vez es más común el uso de equipo de video para realizar este tipo de observaciones. Las videograbadoras tienen la ventaja de proporcionar un registro permanente del comportamiento observado, el cual puede codificarse más tarde. Puede codificar sus observaciones en una carpeta, y un cronómetro puede ser útil para registrar la duración de los hechos. Otra alternativa son los aparatos de registro computarizados, que pueden emplearse para codificar las conductas observadas, así como para obtener información sobre su duración.

Reactividad El segundo aspecto es la **reactividad;** la posibilidad de que la presencia del observador afecte el comportamiento de las personas (véase el capítulo 5). Como se señaló antes, la reactividad puede reducirse por medio de la observación oculta. El uso de ventanas de una sola vista y de cámaras y micrófonos ocultos puede esconder la presencia de un observador. Como alternativa, la reactividad puede reducirse permitiendo que los individuos se acostumbren a la presencia del observador y de cualquier equipo de registro.

Confiabilidad En el capítulo 5 dijimos que la confiabilidad se refiere al grado en que una medida refleja un puntaje verdadero y no el error de medición. Las medidas confiables son estables, consistentes y precisas. Cuando se utiliza la observación sistemática, generalmente participan dos o más observadores para registrar la conducta. La confiabilidad se indica por medio del acuerdo entre los jueces.

Se han reportado niveles de acuerdo muy altos en casi todas las publicaciones de investigaciones realizadas con observación sistemática (generalmente 80 por ciento de acuerdo o más). En algunos programas de investigación a gran escala, en los que participan muchos observadores durante un periodo de años, éstos son entrenados utilizando cintas de video y se verifica el grado de coincidencia de sus observaciones durante el entrenamiento con los de observadores anteriores (véase Bakeman y Gottman, 1986).

Muestreo Finalmente, se debe mencionar el muestreo de las conductas. Para numerosas preguntas de investigación, la obtención de muestras del comportamiento durante un largo periodo ofrece datos más precisos y útiles que las observaciones únicas y breves. Considere un estudio sobre el hábito de ver la Televisión en el hogar (Anderson, Lorch, Field, Collins y Nathan, 1986). Los investigadores deseaban conocer la forma en que los miembros de la familia ven la televisión. Los autores podrían haber estudiado periodos cortos de esta conducía, quizás durante una sola tarde; sin embargo, este tipo de datos puede distorsionarse por tendencias a corto plazo, como la hora del día, un programa en particular o la variación en las actividades familiares, que influyen en la costumbre de ver la televisión. Un mejor método para abordar esta cuestión es observar la conducta a lo largo del tiempo, que es exactamente lo que los investigadores hicieron. Se instalaron grabadoras y cámaras de video en los hogares de 99 familias; el equipo se programó para grabar siempre que el televisor estuviese encendido. Con este método se grabaron casi 5 000 horas de la conducta de ver Televisión. Puesto que la codificación de esta cantidad de datos consumiría mucho tiempo, Anderson y sus colaboradores tomaron una muestra de la conducta de ver televisión cada 55 minutos para hacer el análisis de datos. Entre otras cosas, encontraron que 15 por ciento del tiempo nadie observa el televisor v que esta conducta se incrementa hasta los diez años de edad y después empieza a disminuir.

ESTUDIOS DE CASO

Un **estudio de caso** ofrece la descripción de un individuo. Generalmente éste es una persona, aunque también puede ser un ambiente, como una empresa, una escuela o un vecindario. En ocasiones, a la observación naturalista se le llama estudio de caso y, de hecho, el método de la observación naturalista y del estudio de caso pueden superponerse. Incluimos los estudios de caso en una categoría separada en este capítulo, ya que estos estudios no necesariamente implican la observación naturalista. En su lugar, el estudio de caso puede consistir en la descripción de un paciente, hecha por un psicólogo clínico, o en el relato histórico de un acontecimiento, como el fracaso de un modelo escolar. Una **psicobiografía**es un tipo de estudio de caso en el que un investigador aplica una teoría psicológica para explicar la vida de un individuo que, de manera general, tiene una importancia histórica (véase Elms, 1994). Así, los estudios de caso

pueden utilizar técnicas como la investigación bibliográfica y las entrevistas telefónicas con personas familiarizadas con el caso, pero sin observación directa (véase Yin, 1994).

Según los propósitos de la investigación, el estudio de caso puede presentarla historia del individuo, sus síntomas, sus conductas características, su reacción a las situaciones o sus respuestas a un tratamiento. Generalmente, un estudio de caso se realiza cuando el individuo posee una condición poco común, notoria o especialmente rara. Un estudio de caso famoso es el de un hombre con una habilidad sorprendente para recordar información (Luria, 1968). El hombre, llamado "S", podía recordar largas listas y pasajes con gran facilidad, aparentemente utilizando imágenes mentales para sus habilidades de memoria. Luna también describió algunas de las desventajas de esta aptitud. Por ejemplo, con frecuencia tenía dificultades para concentrarse porque las imágenes mentales aparecían de manera espontánea e interferían con su pensamiento. Otro ejemplo de un estudio de caso sobre el desarrollo del lenguaje implicó a "Genie", una niña que se mantuvo aislada en su habitación, atada a una silla y que nunca habló hasta que fue descubierta a los 13 años y medio de edad (Curtiss, 1977). Obviamente, Genie carecía de habilidades para el lenguaje. Su caso dio a los psicólogos y lingüistas la oportunidad de tratar de enseñarle habilidades de lenguaje y de descubrir cuáles de éstas podía aprender. Aparentemente, Genie fue capaz de adquirir algunas destrezas rudimentarias del lenguaje, como formar oraciones infantiles, pero nunca desarrolló el total de habilidades de lenguaje.

Un ejemplo reciente ilustra la forma en que los individuos con tipos particulares de daño cerebral pueden permitir que los investigadores prueben sus hipótesis (Stone, Cosmides, Tooby, Kroll y Knight, 2002). El individuo en cuestión, "R. M.'\ tenía un daño extenso en el sistema límbico. Los investigadores estaban interesados en estudiar la habilidad para detectar a los individuos tramposos en una relación de intercambio social. El intercambio social es la parte central de nuestras relaciones: una persona brinda bienes o servicios a otra a cambio de algún otro recurso. Stone y sus colaboradores buscaban evidencias de que el intercambio social puede evolucionar en una especie, únicamente cuando existe un mecanismo biológico para detectar a los individuos tramposos; es decir, aquellos que no responden de manera recíproca al no entregar lo que ofrecen. "R. M." completó dos tipos de problemas de razonamiento. Un tipo implicaba la detección de violaciones de las reglas de intercambio social (por ejemplo se debe cubrir un requisito cuando se recibe un beneficio en particular); el otro tipo se enfocaba en reglas no sociales de acción precautoria (por ejemplo, debe usted tomar esta precaución si se involucra en una conducta especialmente peligrosa). Los individuos sin daño cerebral obtienen tan buenos resultados en une como en otro tipo de medidas. Sin embargo, "R. M." tuvo un mal desempeño en los problemas de intercambio social, pero uno bueno en los problemas precautorios, así como en otras medidas generales de habilidades cognoscitivas. Estos hallazgos sustentan la hipótesis de que nuestra capacidad para involucrarnos en relaciones de intercambio social está fundamentada en el desarrollo de un mecanismo biológico que difiere de las habilidades cognoscitivas generales.

Los estudios de caso son valiosos para informarnos acerca de condiciones largas o poco comunes y, por lo tanto, brindan datos únicos sobre algunos fenómenos psicológicos tales como la memoria, el lenguaje o el intercambio social. El conocimiento obtenido por medio de un estudio de caso también puede conducir a la elaboración de hipótesis que pueden probarse a través de otros métodos.

INVESTIGACIÓN DOCUMENTAL

La investigación documental implica el uso de información reunida con anterioridad para resolver preguntas de investigación. El investigador no reúne en realidad los datos originales, sino que analiza datos ya existentes, como las estadísticas que forman parte de registros públicos (por ejemplo, el número de peticiones de divorcio archivadas), reportes de antropólogos, el contenido de las jareas dirigidas a un editor o la información contenida en una base de datos de cómputo. Judd, Smith y Kidder (1991) distinguieron entre tres tipos de datos de investigación documental: los registros estadísticos, los archivos de encuestras y los registros escritos.

Registros estadísticos

Muchas organizaciones públicas y privadas reúnen registros estadísticos. El U. S. Census Bureau (Oficina de censos de Estados Unidos) guarda el conjunto más extenso de registros estadísticos disponibles para que los investigadores realicen Análisis. También hay otros menos obvios, como las estadísticas de salud pública v los registros de puntajes de pruebas, que tienen organizaciones calificadoras reales como el Educational Testing Service.

El béisbol es conocido por la gran cantidad de registros que mantiene sobre prácticamente cualquier aspecto de todos los juegos que se han llevado a cabo. Estas estadísticas están disponibles para cualquiera que desee examinarlas. Reifman, Larrick y Fein (1991) utilizaron dos fuentes de datos documentales, las estadísticas del béisbol y la temperatura diaria, para estudiar la relación entre el calor y la agresión. Estos autores examinaron todos los juegos diurnos de las grandes ligas que se llevaron a cabo en estadios al aire libre durante un periodo de tres años, con el fin de registrar el número de bateadores golpeados por la bola en los lanzamientos. Golpear a un bateador es la medida de la agresión. También revisaron las temperaturas de cada uno de los días de juego. Los resultados mostraron una clara relación entre la temperatura y la agresión, más bateadores fueron golpeados los días en que la temperatura superaba los 90° (Fahrenheit), que cuando estaba en los 80°. La cifra era aún menor cuando la temperatura estaba por debajo de los 80°.

Los registros públicos también pueden emplearse como fuentes de datos documentales. Por ejemplo, Gwaltney-Gibbs (1986) utilizó solicitudes de matrimonio en un condado de Oregón, en 1970 y 1980, para estudiar los patrones

cambiantes de la cohabitación prematrimonial. La autora encontró que en 1970 únicamente el 13 por ciento de las parejas utilizó la misma dirección en la solicitud, pero que el 53 por ciento de las parejas lo hizo en 1980. También pudo relacionar la cohabitación con otras variables como la edad y la raza. Los resultados se interpretaron como sustento de la idea de que la cohabitación prematrimonial se ha convertido en un nuevo paso en los patrones del cortejo que lleva al matrimonio. Otro ejemplo del uso de los registros públicos para el estudio de las relaciones entre la temperatura y la agresión es la investigación de Anderson y Anderson (1984), la cual demostró que existe una relación entre la temperatura y las estadísticas de crímenes violentos en dos ciudades de Estados Unidos.

Archivos de encuestas

Los archivos de encuestas consisten en datos que resultan de consultas almacenadas en computadoras y están disponibles para los investigadores que desean analizarlos. Las principales empresas de encuestas han puesto estos documentos a la disposición del público. Además, muchas universidades forman parte del *Inter-university Consortium for Political and Social Research* (ICPSR), que pone a la disposición datos de archivos de encuestas. Un conjunto de datos muy útil es el *General Social Survey*, una serie de encuestas financiadas por la *National Science Foundation y* que funciona como un recurso para los científicos sociales (Russell y Megaard, 1988). Cada encuesta incluye más de 200 preguntas que cubren un rango de temas como las actitudes, la satisfacción vital, la salud, la religión, la educación, la edad, el género y la raza. Los archivos de encuestas se están volviendo asequibles por medio de Internet, en sitios que permiten que los investigadores analicen los datos en línea. Los archivos de encuestas son sumamente importantes, ya que los investigadores no cuentan con los recursos económicos para realizar sondeos de muestras nacionales seleccionadas aleatoriamente; los archivos les permiten tener acceso a dichas muestras para probar sus ideas.

Registros escritos y de comunicación masiva

Los registros escritos son documentos como diarios y cartas que han sido preservados por sociedades históricas, etnografías de otras culturas escritas por antropólogos y documentos públicos tan diversos como discursos de políticos o mensajes dejados por usuarios de Internet. Los registros de comunicación masiva incluyen libros, artículos de revistas, películas, programas de televisión y periódicos.

Como ejemplo de una investigación documental que utilizó este tipo de registros, Schoeneman y Rubanowitz (1985) estudiaron las cartas de "querida Abby" y "Ann Landers", publicadas en los periódicos. Los investigadores estaban interesados en las causas que las personas asignaban a los problemas que referían en sus cartas. Las cartas se codificaron dependiendo de si los escritores hablaban de sí mismos o de otras personas, y de si las causas comentadas en las

cartas eran de tipo interno (causadas por las propias acciones o personalidad del individuo) o de tipo externo (propiciadas por alguna situación externa a la persona). Cuando las personas hablaban de sí mismas, las causas de los problemas eran principalmente de tipo externo; pero cuando describían a otros individuos, consideraban que los problemas tenían una causa interna (véase también Fischer, Schoeneman y Rubanowitz, 1987).

Los datos documentales también pueden utilizarse en investigaciones de tipo transcultural, con el propósito de examinar aspectos de la estructura social que difieren de una sociedad a otra. Una variable como la presencia *vs.* la ausencia de relaciones matrimoniales monógamas no puede estudiarse en una sola sociedad. En Estados Unidos, por ejemplo, la monogamia es la norma y la bigamia es ilegal. Al estudiar diversas culturas, algunas monógamas y otras no, podemos aumentar nuestros conocimientos sobre las razones por las que se prefiere un sistema o el otro. Este método fue adoptado en* un estudio de Rosenblatt y Cozby (1972) sobre el papel de la libertad de elección de la pareja. Algunas sociedades tienen muchas restricciones con respecto a la pareja que se elige para casarse; otras sociedades ofrecen una gran libertad de elección a las personas jóvenes para que decidan con quién casarse. En el estudio se utilizaron las descripciones de antropólogos (llamadas *etnografías)* de varias sociedades, para calificarlas como de alta o baja libertad de elección de la pareja para casarse. Las etnografías también brindan información sobre muchas otras variables. Los resultados indicaron que cuando existe libertad de elección, el amor romántico y la atracción sexual son bases importantes para seleccionar a la pareja; pero existe un mayor antagonismo en las interacciones entre hombres y mujeres jóvenes. El estudio de Rosenblatt y Cozby utilizó el *Human Relations Área Files* (HRAF), recurso que está disponible en la mayoría de las bibliotecas de universidades para obtener información sobre etnografías. El HRAF consiste en descripciones de muchas culturas, hechas por antropólogos, que se han organizado en categorías como costumbres del cortejo y matrimonio, y prácticas de crianza infantil. Así, es relativamente fácil encontrar información específica de un número mayor de sociedades por medio del HRAF.

Análisis de contenido de documentos

El **análisis de contenido** es el análisis sistemático de documentos existentes como los descritos en esta sección (véase Weber, 1990). Igual que la observación sistemática, el análisis de contenido requiere que los investigadores elaboren sistemas de codificación que puedan ser utilizados por evaluadores para cuantificar la información contenida en los documentos. En ocasiones, la codificación es muy simple y directa; por ejemplo, es fácil codificar si la dirección de la novia y del novio en las solicitudes de matrimonio son iguales o diferentes. Es más común que los investigadores deban definir categorías para codificar la información. En el estudio transcultural de Rosenblatt y Cozby, los evaluadores tenían que leer la información etnográfica y determinar si cada cultura ofrecía una alta o baja libertad de elección de la pareja. Los evaluadores fueron entre-

nados para utilizar el sistema de codificación, y se calcularon coeficientes de confiabilidad entre jueces para asegurar un alto nivel de acuerdo entre éstos. En estudios que examinen archivos documentales como discursos, artículos de revistas, programas de televisión y cartas, deberían utilizarse procedimientos similares.

El empleo de datos documentales permite a los investigadores estudiar preguntas interesantes, algunas de las cuales no podrían estudiarse de otra forma Los datos documentales constituyen un valioso complemento para los métodos más tradicionales de recolección de datos. Sin embargo, existen al menos dos problemas principales en el uso de datos documentales. Primero, a veces es difícil obtener los registros deseados: pueden estar ubicados en almacenes olvidados a tal vez ya hayan sido destruidos. En segundo lugar, nunca podemos estar completamente seguros de la precisión de la información reunida por otra persona

DESCRIPCIÓN DE LA PERSONALIDAD Y DIFERENCIAS INDIVIDUALES

¿Cómo estudiamos la personalidad? ¿Cómo observamos una característica de la personalidad como la extraversión? Como señalamos en el capítulo 4, la extroversión y otras variables de la personalidad son constructos teóricos que no pueden observarse de manera directa. Los investigadores elaboran medidas para evaluar la personalidad. Un tema tan complejo como éste requiere de los métodos cuantitativo y cualitativo. Un área importante de la psicología ha sido la elaboración de medidas cuantitativas de las diferencias individuales de atributos psicológicos como la inteligencia, la autoestima, la extroversión y la depresión. Por ejemplo, Costa y McCrae (1985) crearon el Inventario NEO de la personalidad (NEO-PI) para medir cinco importantes dimensiones de la personalidad: el neuroticismo, la extroversión, la apertura a la experiencia, la sociabilidad y la escrupulosidad. Otras pruebas se enfocan en características muy específicas de la persona, como la "búsqueda de sensaciones", la "ansiedad social" y el "estilo de amar". Incluso, otras medidas se centran en diagnósticos clínicos (por ejemplo, el MMPI-II), en ayudar a las personas a decidir sobre posibles carreras profesionales (por ejemplo, el Inventario de intereses vocacionales) y en la evaluación de individuos que solicitan empleo, por medio de la administración de diversas medidas.

Otra área importante de la investigación de la personalidad estudia la forma en que las características de la personalidad se relacionan con el comportamiento del individuo y sus interacciones con los demás. Por ejemplo, usted podría estudiar la manera en que individuos clasificados como introvertidos o extrovertidos recuerdan el material leído bajo condiciones de distracción contra condiciones de tranquilidad. En el capítulo 8 exploraremos diseños de investigación que abordan temas como éstos.

Los psicólogos estudian muchos atributos que constituyen diferencias individuales, además de la personalidad. Por ejemplo, analizan los intereses y aptitudes relacionados con la *elección* de *carrera, y* el desempeño en el trabajo. El campo de la psicología evaluativa y de la psicometría se enfoca en la elaboración de medidas precisas de las características psicológicas, y en la realización de investigaciones que utilicen las medidas para ayudarnos a comprender mejor la conducta humana.

Generalmente, es mejor emplear medidas existentes de las características psicológicas que desarrollar las propias. Las medidas existentes tienen datos de confiabilidad y validez, y usted puede comparar sus hallazgos con investigaciones anteriores que hayan empleado esta medida. Las fuentes de información sobre pruebas psicológicas elaboradas incluyen el *Mental Measurements Yearbook* (Conoley y Kramer, 1989) y *Test Critiques* (Keyser y Sweetland, 1991). Estos libros de referencia se publican periódicamente y contienen descripciones y evaluaciones de muchas pruebas psicológicas.

Aun cuando la mayoría de los intentos por comprender la personalidad en años recientes son de tipo cuantitativo, los métodos cualitativos para estudiarla representan cada vez más una alternativa. Por ejemplo, los investigadores se están enfocando en explicaciones narrativas de historias vitales para comprender las formas en que se desarrolla la personalidad, y es influida por los sucesos vitales idiosincrásicos, tanto comunes como únicos (Baumeister y Newman, 1994; Josselson y Lieblich, 1993). Probablemente estas explicaciones provengan de fuentes como entrevistas y escritos autobiográficos. Las explicaciones pueden estar relativamente poco enfocadas o pueden estar dirigidas a áreas específicas de la vida, como las relaciones mterpersonales. Más importante, este tipo de análisis cualitativos puede producir datos que difícilmente se obtienen con medidas tradicionales de la personalidad; asimismo, desafían a las teorías tradicionales de la personalidad y conducen a una mayor comprensión del comportamiento humano.

Este capítulo le ha brindado a usted gran cantidad de información sobre importantes métodos cualitativos y cuantitativos de observación, que pueden emplearse para estudiar una gran variedad de preguntas sobre el comportamiento. En el siguiente capítulo exploraremos una forma muy común de examinar el comportamiento humano, sencillamente pidiendo a las personas que utilicen autorreportes para contarnos cosas sobre ellas.

Términos de estudio

Análisis de caso negativo
Análisis de contenido
Estudio de caso
Investigación documental
Observación naturalista
Observación participante
Observación sistemática
Psicobiografía
Reactividad
Sistema de codificación

Preguntas de repaso

1. ¿Que es la observación naturalista? ¸De que manera el investigador reúne datos cuando realiza una investigación con observación naturalista?
2. ¿Por qué los datos de la investigación con observación naturalista son principalmente cualitativos?
3. Indique la diferencia entre la observación participante y la observación ne participante, así como entre la observación encubierta y la observación no encubierta.
4. ¿Qué es la observación sistemática? ¿Por qué los datos de la observación sistemática son principalmente cuantitativos?
5. ¿Qué es un sistema de codificación? ¿Qué aspectos deben considerarse cuando se elabora un sistema de codificación?
6. ¿Qué es un estudio de caso? ¿Cuándo se utilizan los estudios de caso? ¿Qué es una psicobiografía?
7. ¿Qué es la investigación documental? ¿Cuáles son las principales fuentes de datos documentales?
8. ¿Qué es un análisis de contenido?

Actividades

1. Algunas preguntas se responden con mayor facilidad utilizando técnicas cuantitativas, otras, por medio de técnicas cualitativas o una combinación de ambas aproximaciones. Suponga que usted está interesado en la forma en que el alcoholismo de los padres afecta la vida de un adolescente. Elabore una pregunta de investigación que pueda responderse mejor con el uso de técnicas cuantitativas y otra que se ajuste más a las técnicas cualitativas. Una pregunta cuantitativa es: "¿Tienen los adolescentes con padres alcohólicos mayores posibilidades de tener antecedentes penales? y una pregunta cualitativa es: "¿Cómo influyen los padres alcohólicos en las relaciones de sus hijos adolescentes con sus pares?".
2. Elabore un sistema sencillo de codificación para hacer un análisis de contenido de anuncios impresos que aparecen en revistas populares. Aplique el sistema al ejemplar de una revista y describa sus hallazgos.

7

Preguntas hechas a las personas acerca de sí mismas: investigación con encuestas

¿Por qué realizar encuestas?

Elaboración de preguntas

Definición de los objetivos de investigación

Actitudes y creencias

Hechos y datos demográficos

Conductas Redacción de las preguntas

Simplicidad

Preguntas dobles

Preguntas sesgadas

Uso de la negación

Tendencia a "decir sí" y "decir no "

Respuestas a las preguntas

Preguntas cerradas *vs.* preguntas abiertas Número de opciones de respuesta Tipos de escalas

Escala de puntuación gráfica

Escala de diferencial semántico

Escala no verbal para niños Etiquetas para las opciones de respuesta

Finalización del cuestionario

Formato del cuestionario

Refinamiento de las preguntas

Aplicación de encuestas

Cuestionarios

Aplicación personal a grupos o individuos

Encuestas por correo

Encuestas por vía Internet

Otras tecnologías

Entrevistas

Entrevistas cara a cara

Entrevistas vía telefónica

Entrevistas en grupos de enfoque

Diseños de encuestas para estudiar cambios a lo largo del tiempo

Muestreo a partir de una población

Intervalos de confianza

Tamaño de la muestra

Técnicas de muestreo

Muestreo probabilístico

Muestreo aleatorio simple
Muestreo aleatorio por estratos
Muestreo por racimos
Muestreo no probabilístico
Muestreo accidental Muestreo por cuotas
Evaluación de muestras
Encuadre de la muestra
Tasa de respuesta Razones para el uso de muestras por conveniencia
Términos de estudio
Preguntas de repaso
Actividades

La investigación con encuestas emplea cuestionarios y entrevistas para pedir a las personas que proporcionen información sobre ellas, sus actitudes y creencias, datos demográficos (edad, género, nivel de ingresos, estado civil, por ejemplo) y otros hechos, así como comportamientos pasados o futuros. En este capítulo se exploran métodos para el diseño y la realización de encuestas, los cuales incluyen técnicas de muestreo.

¿POR QUÉ REALIZAR ENCUESTAS?

Durante los últimos años se han publicado numerosas historias sobre hallazgos de encuestas en los periódicos. Una de ellas se refería a una encuesta con votantes de Aguascalientes para medir su preferencia entre los candidatos a gobernador. Otra era una encuesta nacional que preguntaba a los adultos sobre sus pensamientos y emociones actuales con respecto de los sucesos del 11 de septiembre del 2001 en Nueva York. Una encuesta con adolescentes reportó sus experiencias con la violencia. También se hacen encuestas en las universidades, por ejemplo para consultar a los egresados sobre su empleo y la percepción que tienen con respecto de su experiencia educativa en el programa que llevaron. En varias tesis de licenciatura y de posgrado se han aplicado cuestionarios a víctimas de violencia familiar. Incluso quienes recientemente han comprado un automóvil nuevo, reciben del fabricante o del distribuidor cuestionarios acerca de su satisfacción con el vendedor y con la calidad del automóvil. Claramente, las encuestas son un método común e importante para estudiar el comportamiento.

Las encuestas nos brindan una metodología para pedir a las personas que nos hablen sobre ellas. Se han vuelto sumamente importantes, ya que la sociedad demanda datos sobre temas y no sólo intuición y anécdotas. También puede ser que se necesiten datos de individuos con posgrado para determinar los cambios que deben hacerse en el curriculum, Las compañías de automóviles obtienen información de los compradores para evaluar y mejorar la calidad de los productos y la satisfacción del cliente. Sin la recolección de estos datos, dependemos por completo de las historias que escuchemos o de las cartas que podría

escribir un individuo con posgrado o un cliente. Otras encuestas pueden ser importantes para que los legisladores y las agencias gubernamentales tomen decisiones sobre políticas públicas. En la investigación básica, muchas variables importantes, entre ellas las actitudes, los estados emocionales *y* los autorre-portes del comportamiento, se pueden estudiar con mayor facilidad por medio de cuestionarios o entrevistas.

Con frecuencia consideramos que los datos de encuestas nos brindan una "imagen instantánea" de la forma en que la gente piensa y se comporta en un cierto momento. Sin embargo, el método de la encuesta también es una herramienta importante para que los investigadores estudien las relaciones entre las variables y la forma en que las actitudes y las conductas cambian a lo largo del tiempo. Por ejemplo, Steinberg y Dornbusch (1991) examinaron la relación entre el número de horas de trabajo de estudiantes de preparatoria con variables como la calificación promedio, el consumo de drogas y de alcohol y los malestares psicosomáticos. La muestra consistió en 3 989 estudiantes de primero a tercer grados, de nueve preparatorias en California y Wisconsin. Los investigadores encontraron que "muchas horas de trabajo durante el año escolar están asociadas con una menor dedicación y desempeño académicos, mayor malestar psicológico y somático, mayor consumo de drogas y alcohol, mayor delincuencia y mayor autonomía de los padres" (Steinberg y Dornbusch, 1991, p. 304). La figura 7.1 presenta un hallazgo típico: hay algunos aspectos positivos de trabajar menos de diez horas a la semana (en oposición a no tener empleo); no obstante, a mayor cantidad de horas laborales se asocian efectos más negativos.

La investigación con encuestas también es importante como complemento de los hallazgos de investigaciones experimentales. Recuerde que en el capítulo 2 comentamos que Winograd y Soloway (1986) realizaron experimentos sobre las situaciones que conducen a olvidar dónde colocamos alguna cosa. Para estudiar este tema por medio de métodos de encuesta, Brown y Rahhal (1994) preguntaron a adultos jóvenes y mayores sobre sus experiencias reales cuando escondían alguna cosa y después olvidaban su localización. Reportaron que a

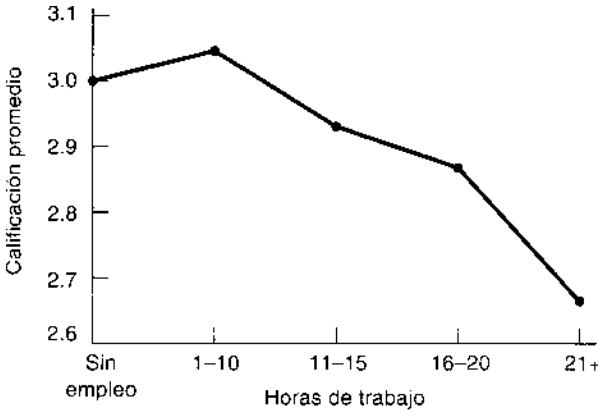


FIGURA 7.1
Relación entre las horas de trabajo y el promedio de las calificaciones escolares. *Fuente:* "Negative Correlaces of Part-time Employment During Adolescente", de L. Steinberg y S. M. Dornbusch, 1991. *Developmental Psychology, 27,* pp. 303-313. Derechos reservados © 1991, Asociación Psicológica Americana. Reproducido baje permiso.

los adultos más grandes les toma más tiempo encontrar el objeto que a los adultos más jóvenes, que los adultos más grandes esconden los objetos de ladrones potenciales, mientras que la gente más joven los esconde de amigos y parientes. De manera interesante, la mayoría de los objetos perdidos son encontrados even-tualmente, casi siempre por accidente y en un lugar donde ya se había buscado. Esta investigación ilustra el señalamiento, hecho en capítulos anteriores, de que se necesitan diversos métodos para comprender cualquier conducta.

Al uso de cuestionarios y entrevistas subyace el supuesto de que las personas desean y pueden proporcionar respuestas verdaderas y precisas. Los investigadores han abordado este tema con el estudio de los posibles sesgos en la forma en que la gente responde. Un **conjunto de respuestas** es la tendencia a contestar todas las preguntas desde una perspectiva en particular, en lugar de proporcionar respuestas que estén relacionadas directamente con las interrogantes. Por lo tanto, los conjuntos de respuestas pueden afectar la utilidad de los datos obtenidos de los autorreportes. El conjunto de respuestas más común es el llamado deseo de aceptación social, o "fingirse bueno", el cual hace que el individuo responda en la forma más comúnmente permitida socialmente; es decir, en la forma en que "la mayoría de la gente" lo hace o de la manera más favorable para la persona. El deseo de aceptación social puede ser un problema en muchas áreas de investigación, aunque tal vez sea más grave cuando la pregunta se refiere a ur. tema sensible como la conducta violenta o agresiva, el abuso de sustancias c las prácticas sexuales. Sin embargo, no se debe asumir que la gente distorsiona consistentemente su información. Jourard (1969) sugirió que las persona.-tienden más a mentir cuando no confían en el investigador; si éste comunica de manera abierta y honesta los propósitos y usos de la investigación, si promete dar retroalimentación sobre los resultados y asegura la confidencialidad, entonces puede esperar, dentro de lo razonable, que los participantes den respuestas honestas.

Ahora estudiaremos los principales aspectos de la investigación con encuetas: la elaboración de preguntas, los métodos para presentarlas y las muestras de quienes participarán en la investigación.

ELABORACIÓN DE PREGUNTAS

Debe ponerse mucha atención cuando se escriben las preguntas para los cuestionarios y las entrevistas. Esta sección describe algunos de los factores más importantes que deben tomarse en cuenta a partir de la elaboración de las preguntas.

Definición de los objetivos de investigación

Cuando se elaboran preguntas para una encuesta, lo primero que el científico debe hacer es determinar explícitamente los objetivos de la investigación: ¿qué desea saber? Las preguntas de la encuesta deben estar vinculadas con las interro-

cantes planteadas en la investigación. Con mucha frecuencia, las encuestas se salen de control cuando los investigadores plantean cualquier pregunta que se les ocurre con relación a un tema, sin considerar con exactitud la información útil que podrían obtener de ella. Este proceso generalmente requiere que el investigador decida qué clase de preguntas plantear. Como se señaló antes, existen tres tipos generales de preguntas de encuesta (Judd, Smith y Kidder, 1991).

Actitudes y creencias Las preguntas sobre actitudes y creencias se enfocan en la forma en que los individuos evalúan y piensan con respecto de las cosas. ¿Debe gastarse más dinero en los servicios de salud mental? ¿Está satisfecho con la forma en que la policía respondió a su llamada? ¿Cómo evalúa a este instructor?

Hechos y datos demográficos Las preguntas sobre hechos piden a las personas que indiquen cosas que saben respecto a sí mismos y a su situación. En la mayoría de los estudios es necesario pedir cierta información de tipo demográfico para describir la muestra adecuadamente; por lo general se pregunta la edad y el genero. De acuerdo con el tema de estudio, podrían incluirse preguntas como el grupo étnico, los ingresos, el estado civil, el puesto en el trabajo y el número de hijos. Obviamente, si usted está interesado en hacer comparaciones entre grupos, como hombres y mujeres, debe pedir la información pertinente sobre la membresía de grupo. Sin embargo, no es práctico plantear estas preguntas si no existen verdaderas razones para utilizar esta información.

La petición de otra información objetiva depende del tema de la encuesta. Cada año, la revista *Consumer Reports {Revista del Consumidor)* pide datos sobre las reparaciones que han necesitado muchos de los productos domésticos, como automóviles y lavadoras. En una encuesta sobre salud y calidad de vida se incluirían preguntas objetivas sobre enfermedades y otro tipo de información médica.

Conductas Otras preguntas de encuesta pueden enfocarse en conductas pasadas o futuras. ¿Cuántas *veces* se ejercitó durante 20 minutos o más la semana pasada? ¿Cuántos hijos planea tener? ¿Alguna vez se ha sentido tan deprimido que se reportó enfermo al trabajo?

Redacción de las preguntas

Es necesario tener mucho cuidado para redactar las mejores preguntas para una encuesta. Los psicólogos cognoscitivistas han identificado varios problemas potenciales en la redacción de las preguntas (véase Graesser, Kennedy, Wiemer-Hastings y Ottati, 1999). Muchos de los problemas surgen por dificultades en la comprensión de la pregunta, como: *a)* vocablos técnicos poco familiares, *h) términos* vagos e imprecisos, *c)* oraciones con una estructura gramatical incorrecta, *d)* frases que sobrecargan la memoria de trabajo, y *e)* preguntas llenas de información distorsionada. A continuación veamos una pregunta que ilustra algunos de los problemas identificados por Graesser y sus colaboradores:

¿Alguna vez su madre, padre, hermanas o hermanos consanguíneos, hijas o hijos han sufrido un ataque cardiaco o un infarto al miocardio?

Existe una sobrecarga a la memoria, debido a la longitud de la pregunta y la necesidad de poner atención a todos esos parientes mientras se lee la pregunta, y el individuo debe preocuparse por dos diagnósticos diferentes con relación a cada uno de sus parientes. Además, el término *infarto al miocardio* puede ser poco familiar para la mayoría de las personas. ¿Cómo se redactan las preguntas para evitar este tipo de problemas? Es importante tomar en cuenta los siguientes aspectos al redactar preguntas.

Simplicidad Las preguntas planteadas en una encuesta deben ser relativamente sencillas. Las personas deben ser capaces de comprenderlas y responderlas con facilidad. Se deben evitar términos técnicos que no puedan comprenderse. No obstante, en ocasiones las preguntas se elaboran de manera un poco más compleja para que se comprendan. Esto sucede de forma general cuando se necesita definir una palabra o describir un aspecto antes de plantear la pregunta. Así, antes de preguntar si alguien aprueba la Proposición X, probablemente usted desee proporcionar una breve descripción del contenido de esta medida de votación.

Preguntas dobles Evite las "preguntas dobles" que cuestionan dos cosas a la vez. Una pregunta como "¿deben las personas de la tercera edad dar más dinero para centros recreativos y programas de ayuda alimentaria?", es difícil de responder, ya que abarca dos actitudes potencialmente muy diferentes. Si le interesan ambos temas, elabore dos preguntas.

Preguntas sesgadas Una pregunta sesgada está redactada de manera que conduce a las personas a responder de cierta forma. Por ejemplo, las preguntas "¿está en favor de eliminar los excesos de presupuesto que se desperdician en las escuelas públicas?", y "¿está en favor de reducir el presupuesto de las escuelas públicas?", probablemente produzcan respuestas diferentes. O considere que los hombres tienen menores posibilidades de decir que han "violado" a alguien, que de decir que "obligaron a alguien a tener relaciones sexuales"; de manera similar, las mujeres son menos propensas a aseverar que han sido violadas que a decir que fueron forzadas a tener relaciones sexuales no deseadas (Koss, 1992). Las preguntas que incluyen términos con carga emocional, como *violación, desperdicio, inmoral, impío o peligroso,* pueden influir en la forma en que las personas responden y, por lo tanto, llevarnos a conclusiones sesgadas.

Uso de la negación Evite redactar preguntas utilizando negaciones. El siguiente reactivo las utiliza: "la ciudad no debe aprobar el refugio propuesto para las mujeres". Manifestarse de acuerdo con esta pregunta significa estar en desacuerde con la propuesta. Esta redacción puede confundir a la gente y dar como resultado

respuestas imprecisas. Un mejor formato sería: "la ciudad debe aprobar el refugio propuesto para las mujeres".

Tendencia a "decir sí" y "decir no" Cuando se plantean varias preguntas sobre un tema, existe la posibilidad de que un individuo emplee un conjunto de respuestas para mostrarse de acuerdo o en desacuerdo con todas las preguntas. Esta tendencia se *conoce* como **"decir sí"** o **"decir no".** El problema es que la persona puede, de hecho, estar expresando un verdadero acuerdo; aunque también puede, simplemente, mostrar su acuerdo con cualquier cosa que usted diga. Una forma para detectar este conjunto de respuestas es redactar las preguntas de manera que no sea posible la manifestación de un acuerdo consistente. Por ejemplo, un estudio sobre los patrones de comunicación familiar podría preguntar a las personas qué tanto están de acuerdo con las siguientes afirmaciones: "los miembros de mi familia pasan mucho tiempo juntos" y "yo paso la mayoría de los fines de semana con amigos". De manera similar, una medida de soledad (por ejemplo, Russell, Peplau y Cutrona, 1980) podría redactar algunas preguntas de tal modo que la manifestación de acuerdo significa que el individuo está solo ("me siento aislado de los demás"), y otras con el significado opuesto, de modo que la manifestación de desacuerdo indique soledad (por ejemplo, "me siento parte de un grupo de amigos"). Aun cuando es posible que algún individuo esté legítimamente de acuerdo con ambos reactivos, el estar de acuerdo o en desacuerdo de manera consistente con un conjunto de preguntas relacionadas, redactadas en formatos estándar e invertidos, es un indicador de que el individuo está siguiendo la tendencia a "decir sí" o "decir no".

Greasser y sus colaboradores, crearon un programa de computadora llamado *QUAID* (Question Understanding Aid), que analiza la redacción de preguntas. Los investigadores pueden poner a prueba sus preguntas en línea, en el sitio de Internet de QUAID (http://mnemosyne.csl.psyc.memphis.edu/quaid).

RESPUESTAS A LAS PREGUNTAS

Preguntas cerradas vs. preguntas abiertas

Las preguntas pueden ser abiertas o cerradas. En las preguntas cerradas se da una cantidad limitada de opciones de respuesta; en las preguntas abiertas, los individuos son libres de responder de la forma que deseen. Así, usted puede preguntar a una persona: "¿qué es lo más importante para preparar a los niños para la vida?", seguido por una lista de respuestas a elegir (una pregunta cerrada), o podría dejar esta pregunta abierta para que la persona dé una respuesta. El uso de preguntas cerradas implica un método más estructurado; son más fáciles de codificar y las *opciones de* respuestas son las mismas para todos. Se requiere más tiempo para clasificar y codificar las respuestas de preguntas abiertas, por lo que son más costosas. En ocasiones no es posible clasificar la respuesta de un individuo, debido a que ésta carece de sentido o porque la persona

no pudo pensar en una respuesta. Aun así, una pregunta abierta puede producir información valiosa sobre lo que la gente piensa. Las preguntas abiertas son más útiles cuando el investigador necesita saber qué piensan las personas y cómo ven el mundo, de manera natural; las preguntas cerradas suelen utilizarse cuando las dimensiones de las variables están definidas adecuadamente.

Schwarz (1999) señala que en ocasiones ambos métodos pueden conducirnos a conclusiones diferentes. Este autor cita los resultados obtenidos de una pregunta de encuesta sobre la preparación de los niños para la vida. Cuando "pensar por sí mismos" era una alternativa en una lista de respuestas, el 62 por ciento eligió esta opción; sin embargo, únicamente el 5 por ciento dio esta respuesta cuando se utilizó el formato abierto. Este hallazgo indica la necesidad de comprender bien el tema cuando se plantean preguntas cerradas.

Número de opciones de respuesta

En las respuestas cerradas existe un número fijo de opciones de respuesta. En las encuestas de opinión pública suele ser suficiente una dicotomía sencilla: "si o no" o "de acuerdo o en desacuerdo". En una investigación más básica, generalmente es preferible tener un número suficiente de opciones para permitir a las personas expresarse; por ejemplo, una escala de cinco o siete puntos que vaya de "completamente de acuerdo a completamente en desacuerdo" o de "muy positivo a muy negativo". Una escala de este tipo podría ser como sigue:

Completamente de acuerdo ___ ___ ___ ___ Completamente en desacuerdo

Tipos de escalas

Las escalas como la anterior son muy comunes en numerosas áreas de investigación. Estas escalas requieren que las personas den juicios de "cantidad" en cierto número de dimensiones; por ejemplo, cantidad de acuerdo, gusto o confianza. Las escalas pueden tener muchos formatos diferentes. El formato que se utilice depende de factores como el tema que se investiga. Quizás la mejor forma para comprender la variedad de formatos sea observar algunos ejemplos. La escala más sencilla y directa presenta a las personas cinco o siete opciones de respuesta que incluyen las etiquetas al inicio y al final de la escala, con el fin de definir los extremos. Por ejemplo,

Se debe exigir que los estudiantes universitarios pasen un examen detallado para graduarse.

Completamente de acuerdo___ ___ ___ ___ Completamente en desacuerdo

¿Qué tan seguro está usted de que el acusado es culpable del intento de homicidio?
Totalmente inseguro ___ ___ ___ ___ Muy seguro

Escala de puntuación gráfica Una escala de puntuación *gráfica* requiere que se ponga una marca a lo largo de una línea continua de cien milímetros, que intuye descripciones en sus extremos.

¿Cómo calificaría la película que acaba de ver?
Nada entretenida _____ Muy entretenida

Después, se coloca una regla sobre la línea para obtener la puntuación en una escala que va de cero a cien.

Escala de diferencial semántico La escala de diferencial semántico es una medida ¿el significado de los conceptos que fue creada por Osgood y sus colaboradores (Osgood, Suci y Tannenbaum, 1957). Los individuos califican cualquier concep-to, ya sean personas, objetos, conductas, ideas, sobre una serie de adjetivos bipo-lares y emplean escalas de siete puntos.

Fumar cigarrillos

___: ___: ___: ___: ___: ___: ___ Malo
___: ___: ___: ___: ___: ___: ___ Débil
___: ___: ___: ___: ___: ___: ___ Pasivo

Bueno
Fuerte
Activo

La investigación sobre el diferencial semántico muestra que prácticamente cualquier cosa puede medirse por medio de esta técnica. Se pueden obtener mediciones sobre cosas específicas (mariguana); lugares (la cafetería); personas (el gobernador, contadores); ideas (aborto, reducción de impuestos), y conductas acudir a la iglesia, uso del transporte público). Un amplio cuerpo de investiga-don indica que los conceptos se califican a lo largo de tres dimensiones básicas: la primera y más importante es la *evaluación* (por ejemplo, adjetivos como bueno-malo, astuto-tonto, amable-cruel); la segunda es la *actividad* (activo-pasivo, lento-rápido, excitable-calmado); y la *tercera* es el *potencial* (débil-fuerte, duro-suave, grande-pequeño).

Escala no verbal para niños Es probable que los niños pequeños no comprendan las escalas que acabamos de describir, aunque son capaces de responder. Por ejemplo, podríamos pedirle a un niño que señale la cara que indica cómo se siente con respecto a un juguete:

Señala la carita que muestra si te gustó el juguete.

Etiquetas para las opciones de respuesta

En los ejemplos incluidos hasta ahora, las escalas sólo tienen etiquetas en los extremos. Quienes las responden infieren el significado de las otras opciones de respuesta. Se trata de un método razonable y los individuos generalmente son capaces de utilizar este tipo de escalas sin dificultad. En ocasiones, los investigadores necesitan proporcionar etiquetas para definir con mayor claridad el significado de cada alternativa. A continuación se presenta una opción muy común a la escala de acuerdo-desacuerdo que se presentó anteriormente:

Completamente de acuerdo	De acuerdo	Indeciso	En desacuerdo	Completamente en desacuerdo

Este tipo de escala asume que la opción central es un punto "neutral" interme-dio entre los extremos. Sin embargo, a veces no es posible o deseable tener un¿ escala perfectamente balanceada. Considere una escala en la que un profesor universitario debe calificar el trabajo de un estudiante o de un programa de pos-grado. Esta escala, en particular, necesita una puntuación comparativa de los estudiantes:

En comparación con otros individuos con estudios de posgrado, ¿cómo calificaría el potencial de éxito de este estudiante?

En el 50% inferior	En el 50% superior	En el 25% superior	En el 10% superior	En el 5% superior

Observe que la mayoría de las opciones exige que los evaluadores den una calificación en términos del 25 por ciento superior. Esto se hace así porque los estudiantes que solicitan entrar a estos programas tienden a ser muy brillantes y a tener una alta motivación, y por ello los profesores los califican de manera fa-vorable. La redacción de las opciones busca obligar a los evaluadores a hacer distinciones más finas entre estudiantes que generalmente son muy buenos.

Es especialmente interesante elaborar las etiquetas de las opciones cuando se indaga acerca de la frecuencia de una conducta. Por ejemplo, usted podría preguntar "¿con qué frecuencia se ejercita durante al menos 20 minutos?". ¿Que tipo de escala debe emplear para que las personas respondan esta pregunta? Podría incluir: 1) nunca, 2) en raras ocasiones, 3) en ocasiones, 4) con frecuen-cia. Estos términos transmiten lo que usted quiere decir, pero son inconsisten-tes. A continuación se incluye otro conjunto de opciones, similares a las des-critas por Schwarz (1999):

_____ menos de dos veces a la semana
_____ dos veces a la semana

_____ cuatro veces a la semana
_____ seis veces a la semana
_____ al menos una vez al día

Una escala diferente podría ser la siguiente:

_____ menos de una vez al mes
_____ una vez al mes
_____ una vez cada dos semanas
_____ una vez a la semana
_____ más de una vez a la semana

A la primera, Schwarz (1999) la denomina escala de alta frecuencia, ya que la mayoría de las opciones indican una alta frecuencia del ejercicio. La otra escala se *conoce* como de baja frecuencia. Este autor señala que-las etiquetas deben elegirse con cuidado porque las personas pueden interpretar el significado de la escala de manera distinta, según las etiquetas que se utilicen. Si usted planteara realmente la pregunta sobre el ejercicio, podría elegir opciones distintas a las que se han descrito aquí. Además, su elección se vería influida por factores *co*-mo el tipo de población que está estudiando. Si estuviera estudiando individuos que hacen mucho ejercicio, con seguridad usaría una escala de mayor frecuenta que si se tratara de personas que generalmente no se ejercitan mucho.

FINALIZACIÓN DEL CUESTIONARIO

Formato del cuestionario

Un cuestionario escrito debe tener una apariencia atractiva y profesional, debe estar escrito con limpieza y sin errores de ortografía. Es necesario que la gente identifique con facilidad las preguntas y las opciones de respuesta. Deje espacio suficiente entre las preguntas para que las personas no se confundan cuando lean el cuestionario. Si incluye una escala con un formato específico, que podría ser de *cinco* opciones, utilícela de manera consistente. Por ejemplo, no cambie de una escala de *cinco* puntos a una de cuatro o siete.

También es adecuado considerar cuidadosamente la secuencia en que planteará las preguntas. En general es mejor colocar primero las preguntas más interesantes e importantes para captar la atención de los individuos y motivarlos a completar la encuesta. Roberson y Sundstrom (1990) obtuvieron el porcentaje más alto de cuestionarios devueltos en una encuesta sobre actitudes de empleados, cuando presentaron primero las preguntas importantes y las demográficas al último. Además, es bueno agrupar las preguntas cuando tratan un tema similar. Si hace esto, logrará que su encuesta tenga una apariencia profesional y las personas la tomarán con seriedad.

Refinamiento de las preguntas

Antes de aplicar la encuesta, es bueno mostrar las preguntas a un pequeño grupo de personas *y* pedirles que "piensen en voz alta" mientras las responden. Los participantes pueden elegirse de la población en estudio, puede tratarse de amigos o colegas que sean capaces de dar respuestas razonables a las preguntas. Para el procedimiento de "pensar en voz alta", necesitará pedirle a los individuos que le digan cómo interpretan cada pregunta y cómo responden ante las opciones de respuesta. Este procedimiento brinda información valiosa que puede utilizarse para mejorar las interrogantes. (La importancia de los estudios piloto como éste se analiza en el capítulo 9.)

APLICACIÓN DE ENCUESTAS

Hay dos maneras de aplicar encuestas. Una es utilizar un cuestionario escrito; los individuos leen las preguntas e indican sus respuestas en un formato. La otra es utilizar un formato de entrevista. Un entrevistador plantea las preguntas y registra las respuestas en una interacción verbal personal. Tanto los cuestionarios como las entrevistas pueden presentarse a los sujetos de diversas formas. Examinemos los diferentes métodos para la aplicación de encuestas.

Cuestionarios

En un cuestionario las preguntas se presentan en un formato escrito y los individuos anotan sus respuestas. Hay varias ventajas en el uso de los cuestionarios Primero, generalmente son menos costosos que las entrevistas; además, permiten que los individuos permanezcan completamente anónimos, ya que no se les pide información que los identifique (por ejemplo, nombre, número del seguro social o número de la licencia de conducir). Sin embargo, los cuestionarios re-quieren que los participantes sean capaces de leer y comprender las preguntas Además, muchas personas consideran aburrido sentarse solos a leer las preguntas y luego responderlas, por lo que puede surgir un problema de motivación Los cuestionarios pueden aplicarse personalmente a grupos o individuos, por medio del correo, vía Internet o con el uso de otras tecnologías.

Aplicación personal a grupos o individuos Con frecuencia, los investigadores pue-den distribuir los cuestionarios a grupos o individuos, que podrían encontrarse en un salón de clase universitario, los padres que acuden a una junta escolar; las personas que acuden a un curso para nuevos empleados o estudiantes que esperan una cita con un consejera. Una ventaja de este método es que se tiene una audiencia cautiva con altas posibilidades de completar el cuestionario, una vez que empiezan a contestarlo. Además, el investigador está presente y las personas pueden plantear sus dudas, en caso necesario.

Encuestas por correo Las encuestas pueden enviarse por correo a las personas, a su hogar o a una empresa. Esta es una forma muy barata de contactar a los individuos que fueron seleccionados para la muestra. No obstante, el formato de correo es inconveniente debido a sus tasas de respuesta potencialmente bajas: es muy fácil guardar el cuestionario y olvidarlo entre las otras tareas que la gente debe hacer en el hogar o en el trabajo. Incluso, si empiezan a contestar el cuestionario, puede suceder cualquier cosa que los distraiga o pueden aburrirse v, sencillamente, tirar el formato a la basura. Algunos métodos que se utilizan para incrementar las tasas de respuesta se describen posteriormente en este capítulo. Otra desventaja es que no hay alguien presente para ayudar a la persona si se confunde o si necesita preguntar alguna cosa.

Encuestas por vía Internet Es muy fácil diseñar un cuestionario para aplicarlo vía Internet. Se pueden redactar preguntas abiertas y cerradas, y presentarlas a los participantes. Una vez que el cuestionario se haya completado, se envían las respuestas de inmediato al investigador. Uno de los primeros problemas por considerar es la manera de seleccionar la muestra. Lo más común es que se incluyan listas de encuestas en buscadores, de modo que quienes están interesados en un tema pueden descubrir que alguien está buscando reunir datos. Algunas de las principales organizaciones de encuestas están construyendo una base de datos de las personas que están interesadas en participar en encuestas. Cada vez que realizan una consulta, seleccionan una muestra de la base de datos v envían una invitación a participar por correo electrónico. Internet también está facilitando la obtención de muestras de individuos con características particulares. Existe todo tipo de grupos especiales de interés en Internet, que incluyen personas con una enfermedad específica, de cierta edad, estado civil o profesión. Utilizan grupos de noticias, discusiones por correo electrónico, anuncios y salones de conversación para intercambiar ideas e información. Los investigadores pueden pedir a las personas que emplean estos recursos que participen como voluntarios de encuestas. Una preocupación con respecto a la recolección de datos en Internet es si los resultados serán similares a lo que podría encontrarse por medio de métodos tradicionales. Aunque la investigación sobre este tema no es extensa, los datos indican que los resultados sí son similares (Krantz, Bailará y Scher, 1997; Stanton, 1998).

Otro problema que tienen los datos de Internet es que existe una ambigüedad con relación a las características de los individuos que proporcionan información para el estudio. Para seguir los lincamientos éticos, el investigador establece que solamente pueden participar personas mayores de 18 años; pero, ¿cómo se controla esto? Las personas también pueden mentir sobre su edad, género o grupo étnico. Sencillamente no sabemos *si* esto representa un gran problema. Sin embargo, en la mayoría de los temas de investigación es poco probable que las personas distorsionen más sus datos en Internet que con cualquier otro método de recolección de datos.

Otras tecnologías Los investigadores están aprovechando las nuevas tecnologías como ayuda para la recolección de datos. Una aplicación interesante es la que se observa en estudios dirigidos a obtener muestras de las conductas y emociones de las personas durante largos periodos. El método más común consiste en pedir a las personas que den explicaciones retrospectivas sobre sus conductas o emociones (por ejemplo, ¿con qué frecuencia se sintió enojado durante la semana anterior?). Con los radiolocalizadores, los teléfonos celulares y otros aparatos inalámbricos de comunicación, es posible establecer contacto con las personas en repetidas ocasiones y pedirles un reporte inmediato de sus actividades y reacciones emocionales. Feldman, Barrett y Barrett (2001) llaman a esto "mues-treo computarizado de experiencias". Las respuestas pueden darse en un cuestionario impreso que se devuelve posteriormente, o puede emplearse algún otro tipo de tecnología, como una serie de preguntas aplicadas por medio de un teléfono por tonos o con un programa que funciona con un asistente digital personal.

Entrevistas

El hecho de que una entrevista incluya la interacción entre personas tiene implicaciones importantes. Primero, la gente suele acceder más a contestar las preguntas ante una persona real que a responder un cuestionario que se le envía por correo. Los buenos entrevistadores suelen ser hábiles para convencer a las personas para que participen; así, las tasas de respuesta tienden a ser más altas cuando se emplean las entrevistas. Es común que el entrevistador y el participante establezcan una relación que motive a la persona a responder todas las preguntas y completar la encuesta. Es más probable que la gente deje preguntas sin contestar en un cuestionario escrito que en una entrevista. Una ventaja importante de la entrevista es que quien la realiza puede aclarar cualquier problema que la persona tenga para comprender las preguntas. Además, el entrevistador puede plantear preguntas de seguimiento si necesita esclarecer algunas respuestas.

Un problema potencial de las entrevistas es el **sesgo del entrevistador.** Este término describe todas las desviaciones que pueden surgir por el hecho de que el entrevistador es un ser humano único que interactúa con un semejante. Por le tanto, un problema potencial es que el entrevistador pueda desorientar sutilmente las respuestas del participante cuando muestra, inadvertidamente, aprobación o desaprobación ante ciertas respuestas. En el caso de que hubiese varios entrevistadores, cada uno podría tener diferentes características (por ejemplo atractivo físico, edad o raza), que podrían influir en la forma en que los participantes responden. Otro problema es que los entrevistadores puedan tener expectativas que los conduzcan a "ver lo que están buscando" en las respuestas ele la gente. Estas perspectivas podrían sesgar sus interpretaciones de las respuestas o conducirlos a insistir hasta obtener la respuesta de ciertos participantes, pero no de otros; por ejemplo, cuestionar más a individuos blancos pero no a las personas de otros grupos raciales, o aplicar pruebas a niños mas no a niñas. E. monitoreo y entrenamiento cuidadoso de los entrevistadores ayuda a limita: estas desviaciones.

Ahora podemos examinar tres métodos para realizar entrevistas: cara a cara, vía telefónica y los grupos de enfoque, también llamados grupos focales.

Entrevistas cara a cara Las entrevistas cara a cara exigen que el entrevistador y el participante se reúnan para llevarla a cabo. Generalmente, el entrevistador viaja a la casa u oficina del participante, aunque en ocasiones éste acude a la oficina del entrevistador. Estas entrevistas tienden a ser bastante costosas y requieren de mucho tiempo, por lo cual suelen utilizarse cuando el tamaño de la muestra es relativamente pequeño y existen claros beneficios de la interacción cara a cara.

Entrevistas vía telefónica Casi todas las entrevistas de encuestas a gran escala se realizan vía telefónica. Las entrevistas por teléfono son menos costosas que las que se realizan cara a cara y permiten una recolección de datos relativamente rápida, debido a que muchos entrevistadores pueden trabajar en la misma consulta al mismo tiempo. Además, las técnicas de encuesta telefónica computarizada disminuyen el costo, pues reducen el trabajo y los gastos del análisis de datos. En un sistema computarizado de entrevista telefónica (SCET), las preguntas del entrevistador aparecen en la pantalla de la computadora y los datos se capturan directamente para su análisis.

Entrevistas en grupos de enfoque Un grupo de enfoque es una entrevista con un conjunto de entre seis y diez individuos, reunidos generalmente durante un periodo de dos a tres horas. En un grupo de enfoque se puede explorar prácticamente cualquier tema. Con frecuencia, los miembros del grupo son seleccionados debido a que poseen un conocimiento o interés particular en el tema. Puesto que el grupo de enfoque necesita personas que dediquen tiempo e incurran en ciertos gastos para viajar al lugar de reunión, generalmente se da un incentivo de tipo económico o un regalo a los participantes.

Las preguntas suelen ser abiertas y se plantean a todo el grupo. Una ventaja de este método es la posibilidad de la interacción: las personas pueden responderse unas a otras y un comentario puede propiciar una diversidad de respuestas. El entrevistador debe ser hábil en el trabajo con grupos, con el fin de facilitar la comunicación y para enfrentar los problemas que puedan surgir, como cuando una o dos personas tratan de dominar la discusión o cuando haya hostilidad entre miembros del grupo. Es común que se grabe la discusión y que después se transcriba. Enseguida, se analizan las cintas y las transcripciones para encontrar temas y áreas de consenso y desacuerdo entre el grupo. Hay ocasiones en que las transcripciones se analizan con un programa de cómputo para buscar ciertas palabras y frases. Los investigadores suelen preferir la organización de al menos dos o tres grupos de discusión sobre cierto tema, con el propósito de asegurarse que la información reunida no es única de un conjunto de personas. Sin embargo, dado que cada grupo de enfoque requiere de mucho tiempo, es costoso y proporciona una gran cantidad de información, los investigadores no utilizan mucho este recurso para estudiar cualquier tema.

DISEÑOS DE ENCUESTAS PARA ESTUDIAR CAMBIOS A LO LARGO DEL TIEMPO

Las encuestas suelen estudiar a las personas en un punto del tiempo. En muchas ocasiones, sin embargo, los investigadores desean hacer comparaciones a lo largo del tiempo. Por ejemplo, un periódico local puede contratar a una empresa para que realice una encuesta anual aleatoria entre los residentes del estado. Dado que las preguntas son las mismas cada año, con el paso del tiempo es posible observar cambios en variables como la satisfacción con el área, las actitudes hacia el sistema escolar y la percepción de los principales problemas que enfrenta el estado. De manera similar, cada año se consulta a una gran cantidad de estudiantes de primer año, en universidades de todo el país, para estudiar cambios en la composición, actitudes y aspiraciones de este grupo (Astin, 1987). Asimismo, los investigadores prueban hipótesis con respecto de la forma en que el comportamiento cambia con el tiempo. Por ejemplo, Sebald (1986) comparó encuestas de adolescentes en 1963, 1976 y 1982. Las preguntas de la encuesta estaban elaboradas para saber a quién piden consejo los adolescentes en una diversidad de temas. El principal hallazgo fue que la búsqueda de consejo de los amigos, más que de los progenitores, se incrementó de 1963 a 1976, pero que esta orientación disminuyó de 1976 a 1982.

Otra forma de estudiar los cambios a través del tiempo es realizando un **estudio de panel,** en el que se consulta a las mismas personas en dos ocasiones. En un estudio de panel de "dos momentos", las personas se estudian en dos puntos del tiempo; en un estudio de panel de "tres momentos", se hacen tres encuestas, y así sucesivamente. Los estudios de panel son especialmente importantes cuando la pregunta de investigación aborda la relación entre una variable en el "momento uno" y otra variable en algún "momento dos". Por ejemplo, Hill, Rubín y Peplau (1976) entrevistaron parejas para estudiar variables como la similitud de sus actitudes. Posteriormente, se entrevistó a las mismas personas para determinar si continuaban en pareja y si así era, qué tan satisfechas se sentían. Los resultados demostraron que la similitud en las actitudes, medida en el momento uno, es un precursor de la duración de la relación.

MUESTREO A PARTIR DE UNA POBLACIÓN

La mayoría de los proyectos de investigación utilizan un **muestreo** de participantes, a partir de una población de interés. La **población** se compone de todos los individuos que le interesan al investigador. Por ejemplo, una población de interés en una gran encuesta de opinión pública podrían ser todos los votantes en Estados Unidos. Esto implica que la población de interés no incluye a las personas menores de 18 años, a los criminales convictos, a los visitantes de otros países y a cualquier otra persona que no pueda ejercer el voto. Usted podría realizar una encuesta en la que su población consistiera en todos los estudiantes de

su universidad. Con tiempo y dinero suficientes, un investigador de encuestas podría lograr comunicarse con cada miembro de la población. Muchos países intentan hacer esto cada diez años, cuando realizan el censo oficial de la población completa. Con una población relativamente pequeña, podría ser fácil para usted estudiarla toda.

No obstante, en la mayoría de los *casos*, estudiar a la población total sería una empresa abrumadora. Por fortuna, esto puede evitarse cuando se selecciona una muestra de la población de interés. Con una muestra adecuada podemos utilizar la información obtenida de los participantes que se incluyeron en ella para conocer con precisión las características de la población como un todo. La teoría estadística nos permite inferir cómo es la población, con base en los datos obtenidos de una muestra (la lógica subyacente se *conoce* como *significación estadística* y se abordará en el capítulo 13).

Intervalos de confianza

Cuando los investigadores *realizan* inferencias sobre poblaciones, lo hacen con cierto grado de confianza. La siguiente es una afirmación que usted podría encontrar cuando lee los resultados de una encuesta: "los resultados de la encuesta tienen una precisión de tres puntos porcentuales, utilizando un nivel de confianza de 95 por ciento". ¿Qué nos dice esto? Suponga que usted les pregunta a algunos estudiantes si prefieren estudiar en su casa o en la escuela; los resultados de la encuesta indican que el 61 por ciento prefiere hacerlo en su casa. Ahora sabe que el valor real de la población probablemente está entre 58 y 64 por ciento. A esto se le llama **intervalo de confianza.** Usted puede tener 95 por ciento de confianza de que el valor verdadero de la población está dentro de este intervalo, alrededor del resultado obtenido de la muestra. Sin embargo, puesto que usted tiene sólo una muestra y no la población completa, su resultado puede ser erróneo. El intervalo de confianza le da información sobre la posible cantidad de error. El término formal es el de *error de muestreo,* aunque quizás usted esté más familiarizado con el término *margen de error.* Recuerde que en el capítulo 5 estudiamos el concepto de error de medición: cuando se mide a un individuo en una variable, la puntuación obtenida puede desviarse de la puntuación verdadera debido al error de medición. De manera similar, cuando se estudia una muestra, el resultado obtenido puede desviarse del valor verdadero de la población, a causa del error de muestreo.

Las encuestas que usted lee con frecuencia en los periódicos y el ejemplo anterior se relacionan con porcentajes. ¿Qué sucede con las preguntas que piden información de tipo cuantitativo? La lógica en este caso es muy similar. Por ejemplo, *si usted* también solicita a los estudiantes que reporten la cantidad de horas y minutos que estudiaron durante el día anterior, podría descubrir que el promedio de tiempo fue de *76* minutos. Entonces, se podría calcular un intervalo de confianza con base en el tamaño de la muestra; por ejemplo, el intervalo de confianza de 95 por ciento es igual a 76 minutos, más o menos diez minutos.

Existen muchas posibilidades de que el valor verdadero de la población esté

dentro del intervalo que va de *66* a 86 minutos. En el capítulo 13 se abordará nuevamente el tema de los intervalos de confianza.

Tamaño de la muestra

Es importante señalar que una muestra grande reduce el tamaño del intervalo de confianza. Aun cuando la extensión del intervalo de confianza está determinada por varios factores, el más importante es el tamaño de la muestra. Las muestras grandes tienen mayores posibilidades de producir datos que reflejen con precisión el valor verdadero de la población. Esta aseveración debe parecerle lógica; una muestra de 200 personas de su escuela debe producir datos más precisos sobre su colegio que una muestra de 25 individuos.

¿Qué tan grande debe ser la muestra? El tamaño de la muestra puede determinarse utilizando una fórmula matemática que toma en cuenta la dimensión del intervalo de confianza y la magnitud de la población en estudio. La tabla 7.1 índica el tamaño que necesita tener la muestra para que el porcentaje de ésta tenga una precisión de más o menos 3, 5 y 10 por ciento, dado un nivel de confianza de 95 por ciento. Observe primero que se necesita un tamaño de muestra mayor para aumentar la precisión. Con una población de 10 mil usted necesita una muestra de 370 para lograr una precisión de 5 por ciento; el tamaño de la muestra se incrementa a 964 para una precisión de 3 por ciento. También es importante señalar que la dimensión de la muestra *no* es un porcentaje constante de la magnitud de la población. Muchas personas creen que un mues-treo adecuado requiere de un cierto porcentaje de la población; estas personas suelen quejarse de los resultados de las encuestas cuando descubren que una consulta realizada en un estado completo se llevó a cabo "únicamente" con 70C o 1 000 personas. Sin embargo, en la tabla, usted puede ver que el tamaño necesario de la muestra no cambia mucho, incluso si el volumen de la población se incrementa de 5 mil a 100 mil o más. Como Fowler (1984) indica, "una muestra de 150 personas describirá a una población de 1 500 o de 15 millones con casi e. mismo grado de precisión. . ." (p. 41).

TABLA 7.1 Tamaño de muestra y precisión en la estimación poblacional (nivel de confianza de 95 por ciento).

	Precisión del estimativo			
Tamaño de la población	± 2%	± 5%	± 10%	
2 000	696	322	92	92
5 000	879	357	94	94
10 000	964	370	95	95
50 000	1045	381	96	96
100 000	1 056	383	96	
Más de 100 000	1 067	384	9683	

Noca: los tamaños muéstrales se calcularon utilizando supuestos conservadores sobre la naturaleza de los verdaderos valores poblacionales.

TÉCNICAS DE MUESTREO

Existen dos técnicas básicas de muestreo de individuos, a partir de una población: el muestreo probabilístico y el no probabilístico. En el **muestro probabilístico** cada miembro de la población tiene una probabilidad específica de ser elegido. Este tipo de muestreo es muy importante cuando se desean hacer aseveraciones muy precisas sobre una población en particular, con base en los resultados de una encuesta. En el **muestreo no probabilístico** se desconoce la probabilidad específica que tiene cualquier miembro en particular de ser elegido. A pesar de que este método no es tan sofisticado como el muestreo probabilístico, veremos que es bastante común y útil en muchas circunstancias.

Muestreo probabilístico

Muestreo aleatorio simple En el **muestreo aleatorio simple,** cada miembro de la población tiene las mismas probabilidades de ser seleccionado en la muestra. Si la población está compuesta de 1 000 miembros, cada uno tiene una posibilidad en mil de ser elegido. Suponga que desea obtener una muestra de los estudiantes que acuden a su escuela. Necesitaría una lista de todos ellos y, a partir de ella, se elegirían estudiantes al azar para formar la muestra.

Cuando realizan entrevistas telefónicas, los investigadores generalmente hacen una lista aleatoria de los números telefónicos, con los prefijos de marcado de las residencias de la ciudad o área que se estudia. Esto produce una mues-:ra aleatoria de la población, ya que la mayoría de las casas tienen teléfono (si muchas personas no lo tuvieran, la muestra estaría sesgada). Con el objetivo de realizar una encuesta, algunas compañías proporcionan a los investigadores lis-ras con números telefónicos, en las que ya se han eliminado los números de empresas y números que las compañías telefónicas no utilizan. Tal vez note que es-re procedimiento da como resultado una muestra aleatoria de hogares y no de individuos. Los investigadores de encuestas utilizan otros procedimientos cuando es importante seleccionar al azar a una persona en su hogar.

Muestreo aleatorio por estratos Un procedimiento un poco más complicado es el **muestreo aleatorio por estratos.** Se divide a la población en subgrupos (o es-Tratos) y después se utilizan técnicas de muestreo aleatorio para seleccionar a los miembros de la muestra de cada estrato. Se puede utilizar cualquier cantidad de dimensiones para dividir a la población, pero la dimensión (o dimensiones) elegidas deben ser relevantes al problema que se estudia. Por ejemplo, en una encuesta sobre las actitudes sexuales se podrían establecer estratos con base en la edad, el género y el grado de educación, ya que estos factores están relacionados con las actitudes sexuales. Una estratificación basada en la estatura o el color de los ojos sería ridícula.

El muestreo aleatorio por estratos tiene la ventaja de asegurar que la muestra refleje, exactamente, la composición numérica de los distintos subgrupos. Este tipo de precisión es particularmente importante cuando algunos subgru-

pos representan porcentajes muy pequeños de la población. Por ejemplo, si 1os afroestadounidenses componen el 5 por ciento de una ciudad de cien mil habitantes, es probable que una muestra aleatoria simple de 100 personas no incluya a ningún afroestadounidense; una muestra aleatoria por estratos incluiría cinco afroestadounidenses elegidos al azar en la población. En la práctica, cuando es importante representar a un grupo pequeño dentro de una población, los investigadores "muestrean en exceso" a ese conjunto para asegurarse de que se consulte a una muestra representativa de dicho grupo; se deberá obtener una muestra lo suficientemente grande para poder hacer inferencias sobre la población. Así, si el *campas* de su universidad tiene una distribución de estudiantes simílar a la de la ciudad descrita anteriormente, y usted desea comparar las actitudes de afroestadounidenses y blancos, necesitará hacer un muestreo con un gran porcentaje de estudiantes afroestadounidenses y sólo un pequeño porcentaje de alum-nos blancos para obtener una cantidad razonable de individuos en cada grupo.

Muestreo por racimos Tal vez usted pensará que debe ser difícil obtener una lista con todos los miembros de una población. ¿Qué pasaría si las autoridades de su escuela deciden que usted no tiene acceso a la lista de todos los estudiantes? ¿Qué pasaría si desea estudiar la población que no tiene una lista de sus miembros, como las personas que trabajan en las agencias de salud del municipio? En estas situaciones se puede utilizar una técnica llamada **muestreo por racimos** En lugar de hacer un muestreo aleatorio a partir de una lista de individuos, el investigador puede identificar "racimos" de personas y después obtener una muestra de éstos. Una vez elegidos los racimos, todos los individuos de cada uno se incluyen en la muestra. Por ejemplo, usted podría realizar la encuesta de estudiantes por medio del muestreo por racimos identificando todas las clases, ya que las clases son los racimos de estudiantes. Después podría sacar una muestra aleatoria de las listas de las clases, y pedir a todos los miembros elegidos que respondan su encuesta (asegúrese, por supuesto, de que ninguno conteste la encuesta dos veces).

Es muy común que el uso del muestreo por racimos requiera de una serie de muestras obtenidas de los racimos más grandes a los más pequeños; es decir, un método "multietapas". Por ejemplo, un investigador interesado en estudiar los centros de salud del municipio, podría primero determinar aleatoriamente una cantidad de estados para sacar la muestra y después seleccionar aleatoriamente municipios de cada uno de los estados elegidos. Luego, el investigador tendría que acudir a los centros de salud en cada uno de estos municipios y estudiar a las personas que trabajan en ellas. Observe que la principal ventaja del muestreo por racimos es que el investigador no necesita obtener la muestra a partir de listas de individuos para lograr una muestra verdaderamente aleatoria de participantes.

Muestreo no prohabilístico

En contraste, las técnicas de muestreo no probabilístico son bastante arbitrarias. Se puede definir una población, pero no se hace demasiado esfuerzo para

asegurar que la muestra represente con precisión a la población. No obstante, entre otras cosas, las muestras no probabilísticas son baratas y convenientes. Dos tipos de muestreo no probabilísticos son el muestreo accidental y el muestreó por cuotas.

Muestreo accidental Una forma de muestreo no probabilístico es el **muestreo accidental,** o "por conveniencia". El muestreo accidental podría considerarse un método de "tómalos donde los encuentres" para obtener participantes. Por lo tanto, se podría seleccionar una muestra de estudiantes de su escuela, de cualquier forma que sea conveniente. Podría pararse frente a la cafetería a las 9 A.M., pedirle a las personas de su clase que se sientan a su alrededor que participen, o bien, visitar un par de salones de clase. Desafortunadamente, este tipo de procedimientos tienden a introducir sesgos en la muestra, de manera que ésta no representa con exactitud a la población de todos los estudiantes. Así, si usted eligió su muestra de los alumnos que llegaron a la cafetería a las 11 A.M., se excluyen aquellos que no frecuentan este lugar; también es probable que no tome en cuenta a aquellos de los turnos vespertino y nocturno. En algunos *campus,* esta muestra diferiría de la población de todos los estudiantes, ya que serían más jóvenes, trabajarían menos horas y tendrían mayores posibilidades de pertenecer a una fraternidad. Desviaciones muéstrales como ésta limitan la posibilidad de utilizar los datos muéstrales para estimar los valores poblacionales verdaderos. Sus resultados no podrían generalizarse a la población, sino que únicamente describirían la muestra sesgada que obtuvo.

Muestreo por cuotas Otra forma de muestreo no probabilístico es el **muestreo por cuotas.** Un investigador que utiliza esta técnica elige una muestra que refleja la composición numérica de diversos subgrupos de la población. Por lo tanto, el muestreo por cuotas es similar al procedimiento del muestreo por estratos, descrita anteriormente; sin embargo, cuando se emplea el muestreo por cuotas no se tiene un muestreo aleatorio. Para ilustrar esto, suponga que desea asegurarse de que su muestra de estudiantes incluye 19 por ciento de alumnos de primer año, 23 por ciento de segundo, 26 por ciento de tercero, 22 por ciento de cuarto y 10 por ciento de estudiantes de posgrado, debido a que éstos son los porcentajes de la población total en las clases. Una técnica de muestreo por cuotas aseguraría estos porcentajes, pero tendría que reunir los datos por medio de técnicas accidentales. Si no tiene una cantidad suficiente de estudiantes de posgrado al ubicarse frente a la cafetería, tal vez podría ir a una clase de posgrado para completar la muestra. Aun cuando el muestreo por cuotas es un poco más sofisticado que el accidental, permanece el problema de que no existen restricciones en la forma de elección de los individuos en los diversos subgrupos. La muestra refleja la composición numérica de la población completa, pero los participantes dentro de cada subgrupo se seleccionan de manera accidental.

EVALUACIÓN DE MUESTRAS

Las muestras deben ser representativas de la población de la que se obtienen. Una muestra sin sesgos es aquella que representa de manera precisa a la población. ¿Cómo se crea una muestra libre de sesgos? Primero se realiza un maestreo aleatorio de una población que contenga a *todos* los individuos que la componen. En segundo lugar, se contacta a *todos* los individuos seleccionados para la muestra y se obtienen respuestas completas de ellos. En raras ocasiones se logran estándares como éstos. Aun cuando se utilice el muestreo aleatorio, pueden surgir desviaciones de dos fuentes: el encuadre de la muestra utilizado y una baja tasa de respuestas. Asimismo, a pesar de que las muestras no probabilísticas poseen más fuentes potenciales de sesgos que las muestras probabilísticas existen muchas razones para su uso y para evaluarlas de forma positiva.

Encuadre de la muestra

El **encuadre de la muestra** es la población *real* de individuos (o racimos), a partir de la cual se selecciona una muestra. Es raro que esto coincida a la perfección con la población de interés; siempre se introduce algún sesgo. Si usted define a su población como "residentes de mi ciudad", el encuadre de la muestra puede ser una lista de números telefónicos que utilizará para comunicarse con los residentes entre las cinco de la tarde y las nueve de la noche. Este encuadre de la muestra excluye a las personas que no tienen teléfono o a aquéllas cuyo horario les impide estar en su casa cuando usted realiza las llamadas. Además, si usted utiliza el directorio telefónico para obtener los números, excluiría a las personas cuyos números no aparecen ahí. En otro ejemplo, suponga que desea saber que piensan los médicos sobre la imagen de su actividad profesional en la televisión Un encuadre muestral razonable serían todos los galenos incluidos en el directorio telefónico. De inmediato, usted podría darse cuenta de que ha limitado su muestra a un área geográfica en particular y, más importante, que también ha limitado la muestra a los facultativos que ejercen la práctica privada; los médicos que trabajan en clínicas y hospitales serían excluidos. Cuando evalúe los resultados de la encuesta, necesita considerar qué tanto se ajusta el encuadre de la muestra a la población de interés. Con frecuencia el sesgo introducido es muy pequeño; sin embargo, podría tener sus consecuencias.

Tasa de respuesta

La tasa de respuesta de una encuesta es, sencillamente, el porcentaje de personas en la muestra que realmente respondió la encuesta. Así, si usted envía por correo 1 000 cuestionarios a una muestra aleatoria de adultos de su comunidad y recibe 500 completos, la tasa de respuesta es de 50 por ciento. La tasa de respuesta es importante, ya que indica la cantidad de sesgo que podría haber en la muestra final de participantes. Los individuos que no responden pueden diferir

de muchas formas de los que sí responden, como la edad, el nivel de ingresos, el estado civil y el grado educativo. A menor tasa de respuestas, mayor es la probabilidad de que dichos sesgos distorsionen los hallazgos y, por lo tanto, limiten la posibilidad de generalizar los resultados a la población de interés.

En general, las encuestas por correo tienen tasas de respuesta más bajas que las consultas telefónicas. No obstante, con ambos métodos se pueden tomar medidas para aumentar dichas tasas. En el caso de las encuestas por correo, se puede enviar una tarjeta o carta con una explicación una semana antes de enviar la encuesta. Los recordatorios de seguimiento, incluso un segundo envío del cuestionario, suelen ser eficaces para incrementar la tasa de respuesta. También es muy útil emplear un sobre de retorno con timbre. Incluso, la apariencia de la página de presentación del cuestionario puede ser importante (Dillman, 2000). En las encuestas telefónicas se puede volver a llamar a las personas que no están en su casa y programar una segunda llamada, en un momento más conveniente, con las personas que no pueden ser entrevistadas en un primer momento. En ocasiones, es necesario utilizar incentivos para incrementar la tasa de respuesta, los cuales pueden incluir dinero en efectivo, un regalo o un certificado de regalo para quienes accedan a participar. Se puede añadir al cuestionario enviado un billete nuevo de un dólar como "agradecimiento". Otro incentivo es la oportunidad de ganar un premio. Finalmente, los investigadores deben intentar convencer a las personas de que los propósitos de la encuesta son importantes y de que su participación será una contribución valiosa.

Razones para el uso de muestras por conveniencia

Gran parte de la investigación en psicología utiliza técnicas de muestreo no probabilístico para obtener participantes en encuestas o experimentos. La ventaja de estas técnicas es que el investigador puede obtener participantes sin gastar mucho dinero y tiempo en la selección de la muestra. Por ejemplo, es común que se elija a los participantes de los estudiantes de cursos introductorios a la psicología. Con frecuencia, a estos alumnos se les pide que participen en estudios que realizan los profesores y sus estudiantes; entonces, los estudiantes de estos cursos pueden decidir en qué investigaciones desean participar.

Incluso, en las investigaciones que no utilizan estudiantes universitarios la muestra suele basarse en la conveniencia más que en la preocupación por obtener una muestra aleatoria. Un investigador puede estudiar a los niños que asisten a una escuela primaria en particular. Tal vez esto se deba a que ha establecido una buena relación con los maestros y administradores, por lo cual es muy fácil obtener el permiso para realizar la investigación. A pesar de que la muestra tiene cierto sesgo cuando incluye únicamente a niños de un vecindario, con ciertas características sociales y económicas, mi colega no está demasiado preocupado,

¿Por qué los científicos no se preocupan más por obtener muestras aleatorias para sus investigaciones? La razón más importante es que la investigación se lleva a cabo para estudiar las relaciones entre variables y no para medir con

precisión valores poblacionales. En la investigación de estudiantes de preparatoria, citada anteriormente (Steinberg y Dornbusch, 1991), la muestra incluyó únicamente a estudiantes de preparatoria de California y Wisconsin. Los datos muéstrales proporcionan información sobre el promedio de horas que estos estudiantes trabajan y el promedio de su consumo de alcohol. Sin embargo, los investigadores no estaban interesados en hacer un cálculo preciso de la cantidad de horas que trabajan los adolescentes de todo el país; estaban más interesados en conocer si el número de horas que los adolescentes trabajan se relaciona con variables, como la calificación promedio y el consumo de alcohol.

Estos temas se explorarán con mayor detalle en el capítulo 14. Por ahora, también es importante reconocer que algunas muestras no probabilísticas son más representativas que otras. La muestra estudiada por Steinberg y Dornbusch parece ser muy representativa de los adolescentes estadounidenses en general, aunque los estudiantes provenían únicamente de dos estados, eran de distintas áreas geográficas y de diversas preparatorias. Los alumnos de cursos introductorios a la psicología son bastante representativos de los estudiantes universitarios en general, y la mayoría de las muestras de estudiantes universitarios representan muy bien a los adultos jóvenes. No existen muchos sesgos obvios, particularmente si se están estudiando procesos psicológicos básicos. Otras muestras pueden ser menos representativas de una población en estudio. Hace poco tiempo, un programa sobre temas públicos de una estación de televisión-pidió a los televidentes que marcaran un número telefónico, o enviaran un correo electrónico, para votar en favor o en contra de una medida de control de armas que estaba siendo evaluada por la legislatura; la noche siguiente, el programa anunció que casi el 90 por ciento de los participantes se opuso a la medida. Los problemas de muestreo son obvios: los grupos opuestos al control de armas pudieron comunicarse de inmediato con sus miembros para animarlos a vota: y no había límites con respecto al número de veces que alguien podía hacerle De hecho, el programa recibió aproximadamente cien veces más votos de los que generalmente recibe cuando hace este tipo de encuestas. Es probable, entonces que esta muestra no fuese representativa de la población de la ciudad, incluso, de la audiencia del programa.

Ahora usted posee una buena cantidad de información acerca de los métodos que se utilizan para obtener datos de la gente. Si usted realiza este tipo de investigación, con frecuencia necesitará diseñar sus propias preguntas, siguiendo los lineamientos descritos en este capítulo, y consultar fuentes como Judd *et al,* (1991); y Converse y Presser (1986). No obstante, también puede adaptar pre-guntas y cuestionarios completos que ya han sido utilizados en investigaciones previas. Por ejemplo, Greenfield (1999) estudió el nuevo fenómeno de la adic-ción a Internet, adaptando preguntas de un gran cuerpo de investigación existente sobre la adicción al juego. Considere utilizar preguntas elaboradas con anterioridad, particularmente si han probado ser útiles en otros estudios (no obstante, asegúrese de no violar derechos de autor). Robinson y sus colaboradores (Robinson, Athanasiou y Head, 1969; Robinson, Rusk y Head, 1968; Robinson, Shaver y Wrightsman, 991) han reunido una variedad de instrumentos

que miden actitudes sociales, políticas y ocupacionales, creados por diversos autores.

En el capítulo 4 señalamos que los métodos de investigación no experimental *y* experimental son necesarios para una plena comprensión del comporta-cuento. Los capítulos anteriores se han enfocado en los métodos no experimentales. En el siguiente capítulo iniciamos una detallada descripción de los diseños de investigación experimentales.

Términos de estudio

Aplicación grupal de encuestas Conjunto de respuestas Conjunto de respuestas con tendencia a "decir sí" y a "decir no" Encuadre de la muestra Encuesta por correo Encuesta por vía Internet Entrevista cara a cara Entrevista telefónica Error de muestreo Escala de alta frecuencia Escala de diferencial semántico Escala de puntuación Escala de puntuación gráfica Estudio de panel Grupo de enfoque intervalo de confianza Investigación de encuestas Muestra aleatoria Muestreo Muestreo accidental (por conveniencia) Muestreo aleatorio por estratos Muestreo aleatorio simple Muestreo no probabilístico Muestreo por cuotas Muestreo por racimos Muestreo probabilístico Población Preguntas abiertas Preguntas cerradas Sesgo del entrevistador Tasa de respuestas

Preguntas de repaso

1. ¿Qué es una encuesta? Describa algunas preguntas de investigación que podría aplicar con una encuesta.
2. Diga cuál es la diferencia entre las técnicas de muestreo probabilístico y no probabilístico. ¿Qué implicaciones tiene cada una?
3. ¿Cuál es la diferencia entre el muestreo accidental y por cuotas?
4. ¿Cuál es la diferencia entre el muestreo aleatorio simple, el muestreo aleatorio por estratos y el muestreo por racimos?
5. ¿Por qué los investigadores que desean probar hipótesis sobre las relaciones entre variables no se preocupan mucho por llevar a cabo un muestreo aleatorio?
6. ¿Cuáles son las ventajas y las desventajas de utilizar cuestionarios en una encuesta, con relación a las entrevistas?

7. Diga cuál es la diferencia entre los siguientes métodos de aplicación de encuestas: el cuestionario, la entrevista e Internet.
8. ¿Cuáles son algunos de los factores que se deben tomar en cuenta cuando se elaboran preguntas para encuestas (entre ellas las preguntas y las opciones de respuesta)?
9. Defina el sesgo del entrevistador.
10. ¿Qué es un conjunto de respuestas con tendencia a la aceptación social?'
11. ¿De qué forma afecta el tamaño de la muestra a la interpretación de los resultados de una encuesta?

Actividades

1. En el estudio de empleo de adolescentes (véase la figura 7.1), Steinberg y Dornbusch (1991) relacionaron un gran número de horas de trabajo con calificaciones promedio más bajas. ¿Puede usted concluir que trabajar más horas *causa* calificaciones más bajas? ¿Por qué sí o por qué no? ¿Cómo podría ampliar el alcance de investigación por medio de un estudio de panel?
2. Elija un tema para una encuesta. Elabore al menos cinco preguntas cerradas que usted crea que deben incluirse en dicha encuesta. Para cada pregunta escriba una versión "correcta" y una "incorrecta". Identifique los elementos que caracterizan sus preguntas incorrectas y diga por qué la versión correcta es mejor.
3. Suponga que desea conocer cuántos libros en una librería han sido escritos por mujeres, por hombres o por ambos (la "librería" puede ser la sección de libros de texto de la biblioteca de su escuela, una librería pequeña o grande, o incluso el conjunto de libros que tiene en casa). Dado que puede haber miles de libros, usted decide tomar una muestra en lugar de revisar cada libro. Describa un posible procedimiento de muestreo usando una técnica no probabilística y otro procedimiento con una técnica probabilística. Ahora comente sobre la forma en la que los resultados pueden diferir según la técnica de muestreo utilizada.

7

Diseño experimental

Confusión y validez interna

Experimentos básicos
Diseño de sólo postest
Diseño de pretest-postest
Ventajas y desventajas de ambos diseños

Asignación de los participantes a las condiciones experimentales

Diseño de grupos independientes

Diseño de medidas repetidas
Ventajas y desventajas del diseño de medidas repetidas Contrabalanceo
Contrabalanceo completo
Cuadrados latinos
Bloques aleatorizados
Intervalo de tiempo entre tratamientos
Elección entre el diseño de grupos independientes y el diseño de medidas repetidas

Diseño de grupos apareados

Diseños de investigación del desarrollo
Método transversal
Método longitudinal
Comparación entre el método transversal y el longitudinal
Método secuencial

Términos de estudio

Preguntas de repaso

Actividades

En el método experimental se controlan todas las variables extrañas. Suponga que desea probar la hipótesis de que el hacinamiento afecta el desempeño cognoscitivo. Para nacerlo, usted puede colocar a un grupo de individuos en una habitación hacinada y a otro en una que no lo esté. Entonces, los participantes de cada grupo realizan la misma tarea cognoscitiva. Ahora, suponga que los sujetos del grupo hacinado no se desempeñan tan bien en la tarea cognoscitiva como las personas del que no está hacinado. ¿Puede atribuirse la diferencia en los puntajes de la prueba a las distintas condiciones de hacinamiento? Sí, *si no existe* ninguna otra diferencia entre los grupos. Sin embargo, ¿qué sucede si al grupo hacinado se le realizó la prueba en una habitación sin ventanas y al grupo sin hacinamiento en una que sí tenía ventanas?; por ejemplo, si estaban en dos salones de clases distintos en una preparatoria. En este caso, sería imposible saber si las bajas puntuaciones de los participantes del grupo hacinado se debieron a esta condición o a la falta de ventanas.

CONFUSIÓN Y VALIDEZ INTERNA

En el capítulo 4 estudiamos que el método experimental tiene la ventaja de permitir una interpretación de los resultados relativamente carente de ambigüedades. El investigador manipula la variable independiente para crear grupos que difieran en los *niveles* de la variable; después compara los grupos en términos de sus puntajes en la variable dependiente. El resto de las variables se mantienen constantes, ya sea por medio del *control experimental* directo o de la *aleatorización.* Si los puntajes de los grupos son diferentes, el investigador puede concluir que la variable independiente causó los resultados, porque la única diferencia entre los grupos es la variable manipulada.

A pesar de que la tarea de diseñar un experimento tiene una lógica elegante y exquisitamente sencilla, usted debe estar consciente de posibles peligros. En el experimento hipotético del hacinamiento que se describió antes, se confunden las variables del hacinamiento y la presencia de ventanas. La variable presencia o no de las ventanas no se mantuvo constante. Una **variable de confusión** es aquella que varía al mismo tiempo que la variable independiente; la confusión ocurre cuando los efectos de dos variables se combinan, de tal manera que no es posible determinar cuál de ellas es la responsable de los efectos observados. Si la variable de las ventanas se mantuviese constante, la presencia o ausencia de ventanas podría afectar el desempeño, pero el efecto de éstas habría sido idéntico en ambas condiciones. Por consiguiente, la presencia de ventanas no sería un factor a tomarse en cuenta cuando se interpreta la diferencia entre el grupo hacinado y el no hacinado.

En resumen, ambas habitaciones en este experimento deberían tener ventanas o no tenerlas. Puesto que una habitación tenía ventanas y la otra no, cualquier diferencia en la variable dependiente (puntajes en las pruebas) no puede atribuirse únicamente a la variable independiente (el hacinamiento). Puede

haber una explicación alterna: la diferencia en los puntajes puede estar causada, al menos en parte, por la variable de las ventanas.

Un diseño experimental adecuado implica eliminar las posibles confusiones que resulten en explicaciones alternas. Un investigador puede afirmar que la variable independiente causó los resultados, únicamente si elimina las explicaciones alternas que compiten con su afirmación. Cuando los resultados de un experimento pueden atribuirse con confianza al efecto de la variable independiente, se dice que posee **validez interna** (véase el capítulo 4). Para lograr validez interna, el investigador debe diseñar y realizar el experimento de tal manera que únicamente la variable independiente pueda ser la causa de los resultados.

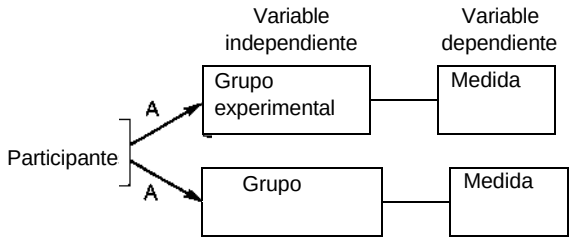
En este capítulo se estudian los diseños experimentales verdaderos, los cuajes proporcionan el nivel más alto de validez interna. En el capítulo 11 examinaremos los diseños cuasi experimentales, que carecen del elemento crucial de la asignación aleatoria, pero que, al mismo tiempo, intentan permitirnos inferir que una variable independiente tuvo un efecto sobre la variable dependiente.

EXPERIMENTOS BÁSICOS

El diseño experimental más simple incluye dos variables: la independiente y la dependiente. La primera posee dos niveles: un grupo experimental y otro de con-trol. Los investigadores deben hacer todo lo posible para asegurarse de que la única diferencia entre los dos grupos sea la variable manipulada. Recuerde, el método experimental implica el control de variables extrañas, ya sea manteniéndolas constantes (control experimental) o utilizando la aleatorización para asegurarse de que cualquier variable extraña afecte a ambos grupos de la misma manera. El diseño experimental básico simple puede tomar una de dos formas: el diseño de sólo postest o el de pretest-postest.

Diseño de sólo postest

Un investigador que utiliza el **diseño de sólo postest** debe: 1) tener dos grupos equivalentes de participantes, 2) introducir la variable independiente y 3) medir los efectos de la variable independiente sobre la dependiente. El diseño es como sigue:



Por lo tanto, el primer paso es elegir a los participantes y asignarlos a los dos grupos. Los procedimientos empleados deben lograr que los grupos sean equi-

valentes para eliminar potenciales **diferencias por selección:** las personas seleccionadas para cada condición no pueden diferir sistemáticamente, de ninguna manera. Para que los grupos sean equivalentes, los individuos se asignan aleatoriamente a las dos condiciones o participan en ambas condiciones. La A en el diagrama significa que los participantes fueron asignados aleatoriamente a los dos grupos.

Después, el investigador debe elegir dos niveles de la variable independiente, como cuando el grupo experimental recibe un tratamiento y el grupo control no. Así, el investigador puede estudiar el efecto de la recompensa sobre la motivación si ofrece un premio a un grupo de niños antes de que practiquen un juego, y no lo hace con los niños del grupo control. Un estudio que prueba los efectos de un método de tratamiento para reducir el tabaquismo, podría comparar a un grupo que recibe el tratamiento con un grupo control que no lo recibe. Otra estrategia sería el uso de dos valores distintos de la variable independiente; es decir, utilizar mayor recompensa en un grupo que en el otro o comparar los efectos de diferentes cantidades de entrenamiento de relajación, diseñado para ayudar a las personas a dejar de fumar (por ejemplo, una hora de entrenamiento comparado con diez horas). Cualquiera de estos métodos proporcionaría una base para comparar a los dos grupos.

Finalmente, se mide el efecto de la variable independiente. En ambos grupos se utiliza el mismo procedimiento de medición, de modo que sea posible compararlos. Cuando los grupos son equivalentes desde el inicio y no existen-variables de confusión, cualquier diferencia entre los grupos con relación a la variable dependiente debe atribuirse al efecto de la variable independiente. El resultado es un diseño experimental con validez interna. En realidad, se podría utilizar una prueba de significación estadística para evaluar la diferencia entre los grupos. Sin embargo, no estamos preocupados por la estadística en este momento. Un experimento debe estar bien diseñado y se deben eliminar las variables de confusión. Si no es así, los resultados son inútiles y la estadística no ayudaría en absoluto.

Diseño de pretest-postest

La única diferencia entre el diseño de sólo postest y el **diseño de pretest-postest,** es que en este último se aplica una prueba antes de introducir la manipulación experimental. Este diseño nos permite estar seguros de que los grupos sean equivalentes al inicio del experimento. No obstante, generalmente no es necesaria esta precaución si los participantes se asignan aleatoriamente a los dos grupos. Con una muestra de participantes de tamaño suficiente, la asignación aleatoria produce grupos que son virtualmente idénticos en todos los aspectos.

Quizás usted se pregunte cuántos participantes se necesitan en cada grupo para asegurarse de que la asignación aleatoria logre grupos equivalentes. A mayor tamaño de la muestra, existen menores posibilidades de que los grupos difieran de manera sistemática antes de la manipulación de la variable independiente. Al mismo tiempo, aumentan las posibilidades de que cualquier dife-

rencia entre los grupos, en la variable dependiente, se deba al efecto de la variable independiente. Existen procedimientos formales para determinar el tamaño necesario de la muestra para detectar un efecto estadísticamente significativo; pero, como regla, usted probablemente necesitará un mínimo de 20 a 30 participantes en cada condición. En algunas áreas de investigación se necesita un número mucho mayor de participantes. En el capítulo 13 se describen con más detalle los aspectos involucrados en la determinación del número de participantes requeridos para un experimento.

Ventajas y desventajas de ambos diseños

Cada diseño tiene ventajas y desventajas que influyen en la decisión de incluir u omitir un pretest. El primer factor de decisión se refiere a la equivalencia de los grupos en el experimento. Aun cuando la aleatorización suele producir grupos equivalentes, es posible que con muestras pequeñas los grupos no sean iguales. Por lo tanto, un pretest permite que el investigador evalúe si los grupos eran realmente equivalentes desde el inicio.

En ocasiones, se necesita un pretest para seleccionar a los participantes del experimento. Un investigador podría aplicar un pretest para encontrar a los individuos que tienen el menor o el mayor puntaje en un cuestionario de tabaquismo, en una prueba de ansiedad ante las matemáticas o en una medición de los prejuicios. Una vez identificados, los participantes se asignan aleatoriamente a los grupos control y experimental. Asimismo, el investigador que utiliza un pretest puede medir la magnitud del cambio en cada individuo. Si un programa de reducción del tabaquismo parece ser eficaz para algunos individuos, pero no para otros, se puede buscar la razón de esto.

También se necesita un pretest siempre que exista la posibilidad de que algunos participantes abandonen el experimento; esto es más probable en un es-:udio que dura mucho tiempo. El factor de abandono en los experimentos se denomina **mortandad.** Las personas pueden abandonar el estudio por razones que no están relacionadas con la manipulación experimental, como en el caso de una enfermedad; en ocasiones, sin embargo, la mortandad está relacionada con la manipulación experimental. Aunque los grupos sean equivalentes al inicio del estudio, las distintas tasas de mortandad pueden convertirlos en grupos no equivalentes. ¿Cómo puede la mortandad afectar un programa de tratamiento diseñado para reducir el tabaquismo? Una posibilidad es que los individuos del grupo experimental que fuman más abandonen el programa. Por lo tanto, ruando se aplique el postest, sólo permanecerán los individuos que fuman poco, de tal manera que la comparación de los grupos experimental y control mostraría menor tabaquismo en el primero; incluso, si el programa no tuviese efecto. De esta manera, la mortandad se convierte en una explicación alterna de los resultados. El uso de un pretest permite evaluar los efectos de la mortandad; se pueden estudiar los puntajes del pretest de los individuos que abandonaron el estudio y saber si la mortandad afectó a los resultados finales. El pretest permite examinar si la mortandad es una explicación alterna posible.

Por lo tanto, el pretest puede ofrecer algunas ventajas en el diseño experimental. Sin embargo, una de sus inconveniencias es que requieren de tiempo v es difícil aplicarlas en el contexto de los procedimientos experimentales particulares que se utilicen. Tal vez lo más importante es que un pretest puede sensibilizar a los participantes al tema que se está estudiando y permitir que descubran sus hipótesis y, así, reaccionarían de manera diferente a la manipulación que s: no hubieran recibido el pretest. Cuando un pretest afecta la forma en que los participantes reaccionan a la manipulación, es muy difícil generalizar los resultados a las personas que no realizaron el pretest; es decir, es probable que la variable independiente no tenga un efecto en la vida real, donde es muy raro que se aplique un pretest. Examinaremos con mayor profundidad este tema en el capítulo 14.

Si la sensibilización por el pretest es un problema, éste puede disfrazarse. Una forma de hacerlo es aplicarlo en una situación completamente diferente, con un experimentador distinto. Otro método implica incluir el pretest en un conjunto de pruebas irrelevantes para que no sea obvio que el investigador esta interesado en un tema en particular.

TABLA 8.1 Diseño de cuatro grupos de Solomon.

	Variable independiente	
Condición de pretest	Grupo control	Grupo experimental
Sin pretest (únicamente postest)		
Pretest y postest		

Nota: si no existe un efecto del pretest, los puntajes promedio del postest en las dos condiciones del grupo control serán iguales, y también las dos medias experimentales del postest. Si existe un efecto del pretest, el patrón de resultados diferirá en las condiciones de sólo postest *y* en la de pretest-postest.

También es posible evaluar de manera directa el impacto del pretest con una combinación de los diseños de sólo postest y de pretest-postest. En este diseño, la mitad de los participantes recibe únicamente el pretest, y la otra mitad recibe el pretest y el postest (véase la tabla 8.1). A esto se le *conoce* formalmente como *diseño de cuatro grupos de Solomon.* Con este diseño, el efecto del pretest puede evaluarse directamente, si el pretest no tiene repercusión, los puntajes del postes: serán iguales en los dos grupos control (con y sin el pretest) y en los dos grupos experimentales. En la figura 8.1 se muestran dos posibles resultados de un diseño de cuatro grupos de Solomon. La primera gráfica ejemplifica un caso en el que los resultados son iguales con el pretest y sin él. La otra gráfica muestra un resultado en el que hay una diferencia entre los grupos experimental y control con la aplicación de un pretest, pero donde no existen diferencias sin éste. El diseño de cuatro grupos de Solomon se estudiará con mayor profundidad en el capítulo 14. Finalmente, se puede utilizar una observación encubierta de la con-

ducta en un pretest, aunque se deben tomar en cuenta los aspectos éticos de hacerlo. Cuando lea sobre un experimento en el que se empleó un diseño de pretest-postest, debe tratar de determinar si el pretest pudo haber provocado un problema al momento de interpretar los resultados.

ASIGNACIÓN DE LOS PARTICIPANTES A LAS CONDICIONES EXPERIMENTALES

Recuerde que existen dos formas básicas de asignación de los participantes a las condiciones. En un procedimiento se asigna a los individuos aleatoriamente a las diversas condiciones, de manera que cada participante pertenece a un solo grupo. A esto se le conoce como **diseño de grupos independientes.** En el otro procedimiento, los individuos participan en todas las condiciones. En el experimento más simple, por ejemplo, cada participante es asignado a ambos niveles de la variable independiente; por esto se le conoce como **diseño de medidas repetidas,** ya que cada participante se mide después de recibir cada nivel de la variable independiente. En las siguientes dos secciones examinaremos con detalle cada uno de estos diseños.

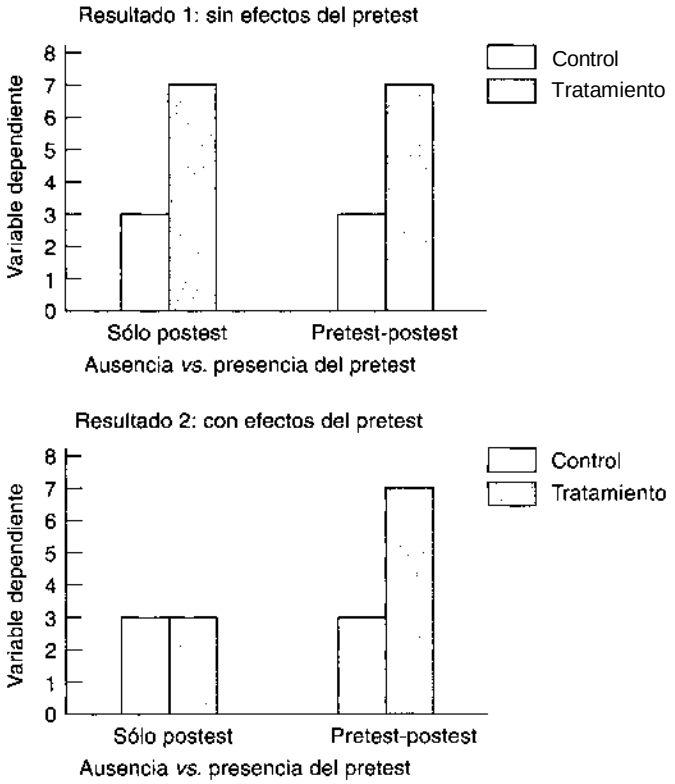


FIGURA 8.1
Ejemplos de resultados del diseño de cuatro grupos de Solomon.

DISEÑO DE GRUPOS INDEPENDIENTES

En un diseño de grupos independientes se asigna a los distintos participantes a cada una de las condiciones por medio de la *asignación aleatoria.* Esto quiere decir que la decisión de asignar a un individuo en una condición particular es completamente al azar y que está fuera del control del investigador. Por ejemplo, usted podría preguntar el mes de nacimiento de los participantes, y asignar en un grupo a aquellos que nacieron en un mes par y en otro a los individuos que nacieron en un mes impar. En la práctica, los investigadores generalmente utilizan una secuencia de números aleatorios para determinar la asignación. En el apéndice C se incluye una tabla de números aleatorios y las instrucciones para utilizarla. La tabla se compone de una serie de números del 0 al 99, que fueron ordenados aleatoriamente por una computadora. El investigador puede utilizar el orden de los números en la tabla para determinar en qué grupo pondrá a cada uno de los participantes. La asignación aleatoria previene sesgos sistemáticos, de modo que todos los grupos son equivalentes en términos de las características de los participantes, tales como el nivel de ingresos, la inteligencia, la edad o las preferencias políticas. De esta forma, las diferencias entre los participantes no pueden constituir una explicación para los resultados del experimento. Como señalamos en el capítulo 4, en un experimento acerca de los efectos del ejercicio sobre la ansiedad, el hecho de que la ansiedad fuese menor en el grupo de ejercicio que en el grupo sin ejercicio, no puede explicarse afirmando que las personas en los grupos diferían de alguna forma en características como el nivel de ingresos, la educación o la personalidad.

Un procedimiento alterno es incluir a los *misinos* individuos en todos los grupos. A éste se le llama diseño experimental de medidas repetidas.

DISEÑO DE MEDIDAS REPETIDAS

Considere un experimento que investiga la relación entre el nivel de significado de un material y su aprendizaje. En un diseño de grupos independientes, a un grupo de participantes se le da material con mucho significado para que lo aprenda, y el otro grupo recibe material con menor significado. En un diseño de medidas repetidas, los mismos individuos participan en ambas condiciones. Así, los participantes primero leerían material con poco significado y responderían una prueba de recuerdo para medir su aprendizaje; después, los mismos participantes leerían material con mucho significado y responderían la prueba de recuerdo. Podrá observar por qué se le llama *diseño de medidas repetidas: se mide* reiteradamente a los participantes en la variable dependiente después de cada condición del experimento.

Ventajas y desventajas del diseño de medidas repetidas

El diseño de medidas repetidas tiene varias ventajas. Una obvia es que se requieren menos participantes, debido que cada individuo participa en todas las condiciones. Cuando hay escasez de participantes o cuando es costoso incluir a cada individuo en el experimento, es preferible un diseño de medidas repetidas. En muchas investigaciones sobre percepción, por ejemplo, se requiere que los participantes reciban un extenso entrenamiento antes de iniciar el experimento verdadero. Este tipo de investigación generalmente incluye, únicamente, pocos individuos que participan en todas las condiciones del experimento.

Otra ventaja del diseño de medidas repetidas es que es sumamente sensible para encontrar diferencias estadísticamente significativas entre los grupos. Esto se debe a que obtenemos datos de las mismas personas en ambas condiciones. Para ilustrar por qué esto es muy importante, considere posibles datos del experimento de recuerdo: con un diseño de grupos independientes, los primeros tres participantes en la condición con material con mucho significado obtuvieron puntajes de 68, 81 y 92. Los primeros tres participantes en la condición con material con poco significado obtuvieron puntajes de 64, 78 y 85. Si usted calcula una puntuación promedio para cada condición, encontrará que el promedio de recuerdo es un poco más alto cuando el material tiene un alto significado. Sin embargo, existe una gran variabilidad en los puntajes de ambos grupos. Ciertamente, usted no está encontrando que todos los individuos en la condición con mucho significado tengan un alto nivel de recuerdo, y que aquellos en la condición con poco significado tengan un bajo nivel de recuerdo. La razón de esta variabilidad es que las personas difieren; hay diferencias individuales en las capacidades de recuerdo, por lo cual existe un rango de puntajes en ambas condiciones. Esto forma parte del "error aleatorio" en los puntajes que no podemos explicar.

No obstante, si los mismos puntajes hubieran sido obtenidos por los tres primeros participantes en un diseño de medidas repetidas, las conclusiones serían muy diferentes. Ordenemos los puntajes de recuerdo para las dos condiciones:

	Alto significado	*Bajo significado*	*Diferencia*
Participante 1	68	64	+4
Participante 2	81	78	+3
Participante 3	92	85	+7

Con un diseño de medidas repetidas, las diferencias individuales pueden observarse y explicarse. Es verdad que algunas personas obtienen mayores puntajes

que otras debido a las diferencias individuales en la capacidad de recuerdo, pero ahora usted puede ver con mayor claridad el efecto de la variable independiente sobre el recuerdo. Es mucho más fácil separar las diferencias individuales sistemáticas de los efectos de la variable independiente: cada participante obtuvo puntajes más altos en la condición con alto significado. Como resultado, tenemos mayores posibilidades de detectar un efecto de la variable independiente sobre la variable dependiente.

El principal problema con el diseño de medidas repetidas proviene del hecho de que las distintas condiciones deben presentarse en una secuencia específica. Suponga que hay mayor recuerdo en la condición con alto significado. Aun cuando este resultado puede ser causado por la manipulación de la variable del significado, el resultado también podría ser sencillamente un **efecto de orden** es decir, que el orden de la presentación de los tratamientos afecte a la variable dependiente. Por lo tanto, un mayor recuerdo en la condición con alto significado podría atribuirse al hecho de que la tarea con este tipo de material ocupa el segundo lugar en el orden de presentación de las condiciones. El desempeño en la segunda tarea podría mejorar, simplemente por la práctica obtenida en la primera tarea.

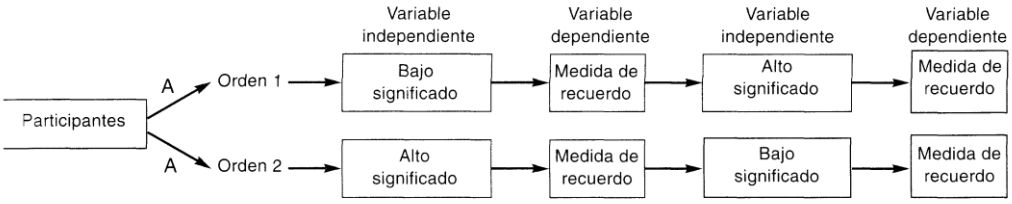
Hay varios tipos de efectos de orden. Los que están asociados con el paso del tiempo incluyen los efectos de práctica y de fatiga. Un *efecto de practica* es una mejoría en el desempeño, como resultado de la práctica repetida de una tarea Un *efecto de fatiga* es el deterioro del desempeño, conforme el participante se cansa, se aburre o se distrae. Los efectos de orden relacionados con el tiempo pueden presentarse siempre que exista una secuencia de tareas por realizar. Por ejemplo, suponga que le pide a un niño que practique un juego de video durante periodos de 30 minutos, bajo distintas condiciones cada vez (por ejemplo, diferentes recompensas por un buen desempeño o diversas cantidades de distracción). El niño que juega por primera vez, puede mostrar un efecto de práctica al obtener mayores puntajes al paso del tiempo, pero un niño que está familiarizado con el juego podría presentar un efecto de fatiga al obtener puntajes cada vez más bajos, cuando se aburra o se canse.

Otros tipos de efectos de orden ocurren cuando el efecto del primer trata-miento influye en la respuesta al segundo tratamiento. Por ejemplo, se presenta un *efecto de contraste* cuando la respuesta a la segunda condición en el experimentó, se ve alterada si las dos condiciones contrastan entre sí. Suponga que la variable independiente es la gravedad de un crimen. Después de leer acerca del crimen menos grave, el más severo podría parecer peor a los participantes de lo que les parecería normalmente. Además, la lectura sobre el crimen más grave podría causar, subsecuentemente, que los participantes consideraran al crimen de menor gravedad mucho más leve de lo que pensarían normalmente.

Existen dos estrategias para enfrentar este tipo de problemas. La primera es utilizar técnicas de contrabalanceo. La segunda es el diseño de un procedimiento en el que el intervalo entre las condiciones sea lo suficientemente largo como para minimizar la influencia de la primera condición sobre la segunda.

Contrabalanceo

Contrabalanceo completo En un diseño de medidas repetidas es muy importante contrabalancear el orden de las condiciones. Con un **contrabalanceo completo** se incluyen todos los órdenes de presentación posibles en el experimento. En el ejemplo del estudio del aprendizaje de material con alto y bajo significado, la mitad de los participantes sería asignada aleatoriamente al orden bajo-alto significado, y la otra mitad al orden alto-bajo significado. Este diseño se ilustra de la siguiente manera:

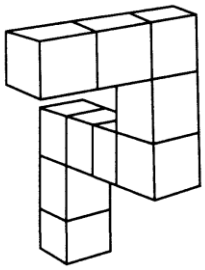


Cuando se contrabalancea el orden de las condiciones es posible determinar el grado en que el orden está influyendo en los resultados. En el estudio hipotético sobre la memoria, sabríamos si el mayor recuerdo en la condición con alto significado es consistente en ambos órdenes de presentación; también sabríamos qué tan responsable de los resultados es el efecto de la práctica.

Los principios del contrabalanceo pueden extenderse a experimentos con tres o más grupos. Con tres grupos existen seis órdenes posibles (3! = 3 x 2 x 1 = 6); con cuatro grupos, el número de órdenes se incrementa a 24 (4! = 4 x 3 x 2 x 1 = 24); se necesitaría un mínimo de 24 participantes para representar cada orden y 48 para tener únicamente dos en cada orden. Imagine el número de órdenes posibles en un experimento realizado por Shepard y Metzler (1971). En su paradigma experimental básico, a cada participante se le muestra un objeto tridimensional, junto con la misma figura girada en uno de diez ángulos diferentes, que van de 0 a 180 grados (véanse los ejemplos de estos objetos en la figura 8.2). En cada ensayo el participante presiona un botón para determinar que las figuras son iguales o diferentes. La variable dependiente es el tiempo de reacción, es decir, la cantidad de tiempo que le toma decidir si las figuras son iguales o discintas. Los resultados muestran que el tiempo de reacción se incrementa conforme el ángulo de rotación se aleja más del original. En este experimento, con diez condiciones, ¡existen 3 millones 628 mil 800 posibles órdenes! Por fortuna, hay alternativas al contrabalanceo completo, que permiten a los investigadores sacar conclusiones válidas acerca de los efectos de la variable independiente.

Cuadrados latinos Una técnica para controlar el orden de los efectos, sin aplicar rodos los órdenes posibles, es la construcción de un **cuadrado latino:** un conjunto limitado de órdenes construido para asegurar que: 1) cada condición apa-

FIGURA 8.2
Figuras tridimensionales.



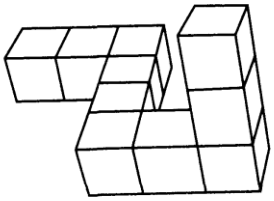


FIGURA 8.3
Cuadrado latino con cuatro condiciones.
Nota: se asignaron letras aleatoriamente a las cuatro condiciones. A = 60 grados, B = 0 grados, C = 180 grados y D = 120 grados. Cada renglón representa un orden distinto de presentación de las condiciones.

	Orden de las condiciones			
	1	2	3	
Renglón 1	**A (60)**	**B (0)**	**D (120)**	**C (180)**
Renglón 2	**B (0)**	**C (180)**	**A (60)**	**D (120)**
Renglón 3	**C (180)**	**D (120)**	**B (0)**	**A (60)**
Renglón 4	**D (120)**	**A (60)**	**C (180)**	**B (0)**

rezca en cada una de las posiciones ordinales y 2) cada condición preceda y siga a cada condición una vez. El uso de un cuadrado latino para determinar el orden, controla la mayoría de los efectos de orden sin tener que incluir todos los órdenes posibles. Suponga que replica el estudio de Shepard y Metzler (1971). utilizando únicamente cuatro de las diez rotaciones: 0, 60, 120 y 180 grados. El cuadrado latino para estas cuatro condiciones se muestra en la figura 8.3. Cada renglón en el cuadrado es uno de los órdenes de las condiciones (éstas tienen 1as etiquetas A, B, C y D). El número de órdenes en el cuadrado latino es igual al número de condiciones; así, si hay cuatro condiciones, hay cuatro órdenes. Cuando usted realice un estudio utilizando el cuadrado latino para determinar el orden necesitará al menos un participante por renglón. Generalmente tendrá dos o más participantes por renglón; el número de ellos en cada orden debe ser el mismo. El procedimiento para la construcción de cuadrados latinos, para cualquier cantidad de condiciones, se incluye en el apéndice D.

Bloques aleatorizados En numerosas áreas de investigación que utilizan diseños de medidas repetidas, el procedimiento experimental básico se repite muchas veces. Por ejemplo, las diez rotaciones del experimento de Shepard y Metzler pueden repetirse varias veces, utilizando figuras originales distintas. A cada re-petición del experimento básico se le llama *bloque de ensayos.* Para controlar los efectos de orden cuando existen muchos de estos bloques de ensayos, se puede determinar aleatoriamente el orden de presentación cada vez.

Debemos señalar que en un experimento en el que se prueba a los individuos durante una serie de ensayos, como sucede en una gran cantidad de estudios del aprendizaje, los "ensayos" constituyen una variable de medidas repetidas. En esta situación, el contrabalanceo no representa un problema; de hecho, al investigador le interesa el efecto de orden de los cambios en el desempeño a lo largo de los ensayos.

Intervalo de tiempo entre tratamientos

Además de contrabalancear el orden de los tratamientos, los investigadores necesitan determinar cuidadosamente el intervalo de tiempo entre la presentación de los procedimientos y las posibles actividades entre ellos. Un periodo de descanso puede contrarrestar el efecto de fatiga; realizar una tarea distinta entre los tratamientos puede reducir la posibilidad del efecto de contraste. Si el tratamiento consiste en la administración de una droga que tarda tiempo en eliminarse, el intervalo entre tratamientos debe ser de un día o más. Wilson, Ellinwood, Mathew v Johnson (1994) examinaron los efectos de tres dosis de mariguana sobre el desempeño de tareas cognoscitivas y motoras. Cada participante fue probado antes y después de fumar un cigarrillo de mariguana. Debido al tiempo que se necesita para que los efectos de la droga se disipen, las tres condiciones se aplicaron en días separados. Con procedimientos que producen cambios emocionales, como un aumento en la ansiedad o el enojo, también se necesitaría un intervalo largo. Quizás haya notado que la introducción de un intervalo extenso puede crear otro problema: los participantes tendrán que comprometerse con el experimento durante más tiempo. Esto puede provocar que sea más difícil reclutar voluntarios y, si el estudio se extiende durante dos días o más, tal vez algunos participantes no regresen.

Elección entre el diseño de grupos independientes y el diseño de medidas repetidas

Los diseños de medidas repetidas tienen dos ventajas principales sobre los diseños de grupos independientes: 1) una reducción en el número de participantes requeridos para completar el experimento y 2) un mayor control de las diferencias entre los participantes y, por lo tanto, una mayor posibilidad de detectar un efecto de la variable independiente. Como se señaló antes, en ciertas áreas de investigación, estas ventajas son muy importantes. Sin embargo, las desventajas de los diseños de medidas repetidas y la necesidad de tomar precauciones para enfrentarlas, suelen ser razones suficientes para utilizar los diseños de grupos independientes.

Una consideración muy diferente para decidir el uso del diseño de medidas repetidas se refiere a la generalización de las condiciones a la "vida real". Green-vald (1976) ha señalado que en las situaciones cotidianas reales, en ocasiones hallamos variables independientes en forma de grupos independientes: nos

encontramos sólo una condición sin otra con la cual compararla. No obstante, algunas variables independientes se encuentran con mayor frecuencia en forma de medidas repetidas: ambas condiciones aparecen y nuestras respuestas ocurren en el contexto de la exposición en ambos niveles de la variable independiente. Por ejemplo, si usted está interesado en la manera en que las características de un acusado afectan al jurado, un diseño de grupos independientes sería más apropiado porque los jueces reales en un juicio se concentran en un solo inculpado. Sin embargo, si se interesa en los efectos que tienen las características de un solicitante de trabajo en el empleador, sería razonable utilizar un diseño de medidas repetidas, debido a que los empleadores suelen considerar a varios solicitantes al mismo tiempo. El uso de un diseño de grupos independientes o de un diseño de medidas repetidas debe estar determinado, en parte, por estos aspectos de la generalización.

Finalmente, cualquier procedimiento experimental que produzca un cambio relativamente permanente en un individuo, no puede emplearse en un diseño de medidas repetidas. Algunos ejemplos incluyen un procedimiento quirúrgico, como las lesiones cerebrales, o un tratamiento psicoterapéutico.

DISEÑO DE GRUPOS APAREADOS

Un método un poco más complicado para la asignación de participantes a las condiciones de un experimento es el **diseño de grupos apareados.** En lugar de sólo asignar aleatoriamente a los participantes en los grupos, la meta es igualar o aparear primero a los participantes, con base en una de sus características. La variable de apareamiento es la medida dependiente o una variable que esté íntimamente relacionada con la variable dependiente. Por ejemplo, en un experimento de aprendizaje, los participantes podrían aparearse con base en los puntajes en una prueba de una habilidad cognoscitiva; pero si la habilidad cognoscitiva no está relacionada con la medida dependiente, el apareamiento sería una pérdida de tiempo. La meta es lograr la misma equivalencia de los grupos que se obtiene por medio del diseño de medidas repetidas, sin la necesidad de que los mismos individuos participen en ambas condiciones.

Cuando se utiliza un diseño de grupos apareados, el primer paso consiste en obtener una medida de la variable de apareamiento de cada individuo. Entonces, se ordena a los participantes del más alto al más bajo con base en sus puntajes en la variable de apareamiento. Ahora el investigador puede formar pares que sean aproximadamente iguales en esa característica (los dos participantes con puntajes más altos forman el primer par, los siguientes dos forman el segundo y así, sucesivamente). Finalmente, los miembros de cada par se asignan aleatoriamente a las condiciones del experimento.

El método de grupos apareados asegura que los grupos sean equivalentes (en la variable de apareamiento), antes de introducir la manipulación de la variable independiente. Esto puede ser especialmente importante con muestras pequeñas, porque los procedimientos de asignación aleatoria tienen mayores

rosibilidades de producir grupos equivalentes cuando aumenta el tamaño de la muestra. El apareamiento, entonces, es más utilizado cuando se dispone de poros participantes o cuando es muy costoso incluir una gran cantidad de individuos en el experimento.

Estas ventajas dan como resultado una mayor posibilidad de encontrar un efecto estadísticamente significativo de la variable independiente, ya que se pueden explicar las diferencias individuales en respuesta a la variable independiente, tal como sucede con un diseño de medidas repetidas.

Estos aspectos de la variabilidad y la significación estadística se estudian con mayor profundidad en el capítulo 13 y el apéndice B. Lo más importante aquí es que aparear con base en una variable incrementa las posibilidades de que se encuentre una diferencia estadísticamente significativa en un experimento. No obstante, los procedimientos de apareamiento pueden ser costosos y tomar mucho tiempo, debido a que requieren que se mida a los participantes en la variable de apareamiento antes del experimento. Estos esfuerzos valen la pena únicamente cuando dicha variable de apareamiento está íntimamente relacionada con la variable dependiente, y cuando se sabe que la relación existe antes de realizar el estudio. Por estas razones, el diseño de grupos apareados no es de uso común.

Una técnica intermedia evita los problemas del apareamiento y al mismo tiempo mantiene algunas de sus ventajas estadísticas. En lugar de obtener de antemano datos de la variable de sujeto (por ejemplo, puntajes de inteligencia), para establecer pares de individuos, se utiliza el diseño de asignación aleatoria, pero se dedica un tiempo para reunir datos de la variable de sujeto. A pesar de que estos datos se reúnen "después del hecho", pueden emplearse para el análisis estadístico de los resultados. Tal vez usted haya visto investigaciones que reportan el uso del "análisis de covarianza", o que se refieren a una variable, como la inteligencia, como una "covarianza". Aquí, una técnica estadística llamada **análisis de covarianza,** controla estadísticamente la correlación entre la variable de los sujetos y la variable dependiente del experimento. Este procedimiento elimina la varianza de error que resulta del hecho de que la variabilidad en los puntajes de la variable dependiente se deba, en parte, al efecto de la variable del sujeto. La teoría y los procedimientos del análisis de covarianza rebasan el alcance de este libro. Sin embargo, una vez que se haya familiarizado con los métodos estadísticos más sencillos descritos en el capítulo 12 y el apéndice B, estará listo para emplear técnicas como ésta.

DISEÑOS DE INVESTIGACIÓN DEL DESARROLLO

Con frecuencia, los psicólogos que estudian el desarrollo investigan la forma en que los individuos cambian en función de la edad. Un investigador puede someter a prueba una teoría sobre los cambios en la capacidad del razonamiento de los niños conforme crecen, la edad en que se desarrolla la conciencia de sí mismos en los niños pequeños o los valores globales de la gente cuando pasa de la adolescencia a la adultez. En todos los casos, la variable principal es la edad.

Método transversal

	Año de nacimiento (cohorte)	Momento 1: 2000
Grupo 1:	1945	55 años
Grupo 2:	1940	60 años
Grupo 3:	1935	65 años

Método longitudinal

	Año de nacimiento (cohorte)	Momento 1: 2000	Momento 2: 2005	Momento 3: 2010
Grupo 1:	1945	55 años →	60 años →	65 años

Método secuencial

	Año de nacimiento (cohorte)	Momento 1: 2000	Momento 2: 2005	Momento 3: 2010
Grupo 1:	1945	55 años →	60 años →	65 años
Grupo 2:	1935	65 años →	70 años →	75 años

FIGURA 8.4
Tres diseños para la investigación del desarrollo.

Los investigadores del desarrollo se enfrentan a una decisión interesante cuando diseñan sus estudios, ya que existen dos métodos generales para el estudio de individuos de distintas edades: el transversal y el longitudinal. Usted verá que el método transversal comparte similitudes con el diseño de grupos independientes, mientras que el longitudinal es similar al diseño de medidas repetidas. También examinaremos un método híbrido denominado secuencial. Las tres aproximaciones se ilustran en la figura 8.4.

Método transversal

En una investigación que utiliza el **método transversal** se estudian personas de diferentes edades en un punto del tiempo. Suponga que está interesado en indagar cómo cambia la capacidad de aprender una aplicación de cómputo conforme las personas crecen. Con el método transversal, podría estudiar a personas que actualmente tienen 20, 30, 40 y 50 años de edad. Los participantes en su estudio realizarían la misma tarea de aprendizaje de cómputo, y usted compararía el desempeño de los grupos.

Método longitudinal

En el **método longitudinal** se observa al mismo grupo de personas, en distintos puntos del tiempo, conforme crecen. Quizás el estudio longitudinal más famoso sea el Terman Life Cycle Study (Estudio de ciclos de vida de Terman), que inició el psicólogo Lewis Terman en Stanford, en 1921. Terman estudió a 1 528 niños de California en edad escolar, con puntajes de cociente intelectual de al menos 135. Los participantes, quienes se denominaron a sí mismos "termitas", fueron medidos inicialmente en diversos aspectos de su desarrollo cognoscitivo y social, en 1921 y 1922. Terman y sus colegas continuaron estudiando a los "termitas" durante su infancia y adolescencia, y también a lo largo de su vida adulta (véase Terman, 1925; Terman y Oden, 1947,1959). Los sucesores de Terman continúan estudiando a los "termitas" hasta su muerte. El estudio ha proporcionado una rica descripción de la vida de individuos muy inteligentes, y ha desmentido muchos estereotipos negativos de los altos niveles de inteligencia; por ejemplo, los "termitas" tenían un ajuste adecuado tanto social como emocional. Se tienen archivados los datos para que los utilicen otros investigadores. Fried-man y sus colaboradores (1995) utilizaron información de Terman para estudiar factores sociales y prácticas de salud, asociados con la edad y con la muerte. Un hallazgo intrigante fue que la dimensión de la personalidad "escrupulosidad" está relacionada con la longevidad.

La mayoría de las investigaciones longitudinales no alcanzan la magnitud del estudio de Terman, en donde se estudió a los individuos durante toda su vida, pero cualquier estudio que intenta comparar a los mismos individuos en diferentes edades está utilizando el método longitudinal. Caspi, por ejemplo, realizó un estudio longitudinal de 15 años, en el que probó a niños neozelandeses a los tres y 18 años de edad. Entre los muchos hallazgos de este estudio, uno mostró que el temperamento a los tres años está relacionado con rasgos de la personalidad a los 18; por ejemplo, los niños "descontrolados" se convirtieron en adultos impulsivos y agresivos (Caspi y Silva, 1995). Otras investigaciones longitudinales suelen estudiar periodos más cortos; por ejemplo, un estudio de ocho años que demostró los efectos positivos de los centros de cuidado infantil en una muestra de niños suecos (Broberg, Wessels, Lamb y Hwang, 1997).

Comparación entre el método transversal y el longitudinal

El método transversal es más común que el método longitudinal, principalmente porque es menos costoso y produce resultados útiles de manera inmediata. Observe que con un diseño longitudinal tomaría tres décadas estudiar al mismo grupo de individuos desde los 20 hasta los 50 años; pero con un diseño transversal se pueden obtener comparaciones entre distintos grupos de edad relativamente rápido.

No obstante, existen algunas desventajas de los diseños transversales. Lo más importante es que los investigadores deben inferir que las diferencias entre los grupos de edad se deben a la variable del desarrollo. El cambio en el desarrollo no se observa directamente entre el mismo grupo de personas, sino que está basado en comparaciones entre distintas cohortes de individuos. Una **cohorte** se define como un grupo de personas que nacieron casi al mismo tiempo, que estuvieron expuestas a los mismos sucesos en una sociedad e influidos por las mismas tendencias demográficas, como las tasas de divorcio y el tamaño de la familia. Si piensa en los peinados de personas que usted *conoce* que tienen 30, 40, 50 y 60 años de edad, de inmediato reconoce la importancia de los efectos de cohorte. Lo más importante es que las diferencias entre las cohortes reflejan distintas condiciones económicas y políticas en la sociedad; diferente música y arte; diversos sistemas educativos y prácticas de crianza infantil. En un estudio transversal, la diferencia entre grupos de distintas edades puede reflejar cambios del desarrollo por la edad; sin embargo, las diferencias pueden provenir de efectos de cohorte (Schaie, 1986).

Para ilustrar esto, regresemos a nuestro estudio hipotético sobre el aprendizaje del uso de las computadoras. Suponga que usted descubre que la edad está relacionada con una disminución en las habilidades, de modo que las personas del grupo de 50 años obtienen menores puntajes en la medida de aprendizaje que las que tienen 40, y así sucesivamente. ¿Debe concluir que la capacidad para aprender a utilizar una computadora disminuye con la edad? Ésta puede ser una conclusión exacta; de manera alternativa, las diferencias podrían deberse a un *efecto de* cohorte: las personas mayores tuvieron menor experiencia con las computadoras durante su crecimiento. El punto importante es que el método transversal confunde la edad con los efectos de cohorte. (Revise el tema sobre la confusión y la validez interna al inicio de este capítulo.) Finalmente, usted debe notar que los efectos de cohorte suelen convertirse en un problema cuando el investigador está estudiando los efectos del tiempo a lo largo de un amplio rango de edades (por ejemplo, de los adolescentes a los adultos mayores).

La única forma de estudiar de manera concluyente los cambios que ocurren conforme la gente crece, es por medio de un diseño longitudinal. A su vez, la investigación longitudinal es la mejor manera para estudiar cómo se relacionan los puntajes en una variable a cierta edad, con otra variable a una edad mayor. Por ejemplo, si un investigador desea estudiar la relación entre el ambiente en el hogar de un niño de cinco años, con el aprovechamiento escolar a los 13, un estudio longitudinal proporciona los mejores datos. La alternativa en este caso sería estudiar a niños de 13 años y preguntarles, a ellos o a sus padres, cómo era el ambiente en su hogar; este método *retrospectivo* tiene sus propios problemas cuando uno considera la dificultad de recordar los acontecimientos del pasado distante.

Por lo tanto, el método longitudinal, a pesar de ser costoso y complejo, posee ventajas definitivas. Pero existe un problema importante: durante el transcurso de un estudio longitudinal, las personas pueden cambiarse de casa, morir o perder el interés en el estudio. Los investigadores que realizan estudios longi-

rudinales tratan de convencer a las personas para que continúen; suelen viajar a cualquier lugar para reunir más datos, y comparan los puntajes de las pruebas de los individuos que abandonan el estudio con aquellos que permanecen, con el fin de hacer mejores análisis de sus resultados. ¡En resumen, un investigador no debe embarcarse en un estudio longitudinal sin una gran cantidad de recursos, paciencia y energía!

Método secuencial

Una alternativa a los métodos longitudinal y transversal es el **método secuencial.** En la figura 8.4 se le ilustra, junto con el transversal y el longitudinal. En la figura, la meta del estudio es comparar individuos de 55 y 65 años de edad. La primera fase del método secuencial se inicia con el método transversal; por ejemplo, se podrían estudiar grupos de individuos de 55 y de 65 años, y después se utilizaría el método longitudinal al probar a cada persona al menos una vez más. Este diseño requiere de un menor número de años para completarse que un estudio longitudinal, y el investigador obtiene recompensas inmediatas porque los datos de los diferentes grupos de edades están disponibles desde el primer año del estudio.

Ahora usted comprende los fundamentos de los diseños de experimentos. En el siguiente capítulo abordaremos temas que surgen cuando se decide cómo llevar a cabo realmente un experimento.

Términos de estudio

Análisis de covarianza
Asignación aleatoria
Bloques aleatorizados
Cohorte
Contrabalanceo
Cuadrado latino
Diferencias de selección
Diseño de grupos apareados
Diseño de grupos independientes
Diseño de medidas repetidas
Diseño de pretest-postest
Diseño de sólo postest
Efecto de contraste
Efecto de fatiga Efecto de práctica Efectos de orden Método longitudinal Método secuencial Método transversal Mortandad
Validez interna Variable de confusión

Preguntas de repaso

1. ¿Qué es la confusión de variables?
2. ¿A qué se refiere la validez interna de un experimento?
3. ¿De qué forma los dos diseños experimentales verdaderos eliminan el problema de las diferencias de selección?

4. Describa las diferencias entre el diseño de sólo postest y el diseño de pretest-postest. ¿Cuáles son las ventajas y desventajas de cada uno de ellos?
5. ¿Cuándo decide un investigador utilizar el diseño de grupos apareados? ¿Cuáles son las ventajas de este diseño?
6. ¿Qué es un diseño de medidas repetidas? ¿Cuáles son las ventajas de utilizar un diseño de medidas repetidas? ¿Cuáles son las desventajas?
7. ¿Cuáles son algunas de las formas que se utilizan para resolver los problemas del diseño de medidas repetidas?
8. El procedimiento empleado para obtener una muestra (por ejemplo, muestreo aleatorio o no aleatorio) no es igual al procedimiento utilizado para asignar a los participantes a las distintas condiciones; establezca la diferencia entre el muestreo aleatorio y la asignación aleatoria.
9. ¿Cuál es la diferencia entre los métodos longitudinal, transversal y secuencial?
10. ¿Qué es un efecto de cohorte?

Actividades

1. Diseñe un experimento para probar la hipótesis de que las clases de matemáticas que sólo incluyen mujeres son benéficas para las adolescentes. Defina operacionalmente las variables independiente y dependiente. Su experimento debe incluir dos grupos y emplear el procedimiento de grupos apareados. Seleccione bien la variable de apareamiento. Además, defienda su elección de un diseño de sólo postest o de un diseño de pretest-postest.
2. Diseñe un experimento de medidas repetidas que investigue los efectos del estilo de presentación de un reporte sobre la calificación obtenida en éste. Utilice dos niveles de la variable independiente: una presentación con "estilo profesional" (papel de alta calidad, uso consistente de márgenes y tipos de letra, tablas y gráficas construidas cuidadosamente) y un "estilo no profesional" (papel de calidad promedio, cambios frecuentes en los márgenes y tipos de letra, tablas y gráficas sin leyendas apropiadas). Hable de la necesidad de utilizar el contrabalanceo. Elabore una tabla que describa el diseño experimental.
3. El profesor Foley realizó una prueba del sabor de refrescos de cola. Cada participante en el experimento primero probó dos onzas de *Coca-Cola,* luego dos onzas de *Pepsicola* y, finalmente, dos onzas de *RC Cola.* Después de cada prueba los individuos calificaban el sabor de los refrescos. ¿Cuáles son los problemas potenciales de este diseño experimental y de los procedimientos utilizados? Revise el diseño y los procedimientos para localizar estos problemas. Tal vez quiera considerar varias alternativas y pensar sobre las ventajas y desventajas de cada una.

4. El doctor Cárdenas estudió las preferencias políticas de diferentes grupos de individuos de 20, 40 y 60 años de edad. Encontró que el grupo de 60 años tenía las preferencias políticas más conservadoras y el grupo de 20 las menos conservadoras.
   a) ¿Qué tipo de métodos utilizó en este estudio?
   b) ¿Podría concluir que las personas se vuelven políticamente más conservadoras conforme se hacen mayores? ¿Por qué sí o por qué no?
   c) Proponga formas alternas para estudiar este tema.

9

Realización de experimentos

Selección de los participantes de la investigación

Manipulación de la variable independiente
Preparación del terreno
Tipos de manipulación
Manipulación directa
Manipulación por etapas
Intensidad de la manipulación Costo de la manipulación

Medición de la variable dependiente
Tipos de medidas
Medidas de autorreporte
Medidas conductuales
Medidas fisiológicas Sensibilidad de la variable dependiente Medidas múltiples Costo de las medidas Ética

Controles adicionales
Control de las expectativas de los participantes
Características de la demanda
Grupos placebo Control de las expectativas del investigador
Investigación sobre los efectos de las expectativas
Soluciones al problema de las expectativas

Consideraciones adicionales
Propuestas de investigación
Estudios piloto
Verificación de la manipulación
Informe

Análisis e interpretación de resultados

Comunicar a otros la investigación
Congresos profesionales
Artículos científicos

Términos de estudio

Preguntas de repaso

Actividades

Los capítulos anteriores establecieron las bases para planear una investigación. En este capítulo nos enfocamos en algunos aspectos muy prácticos de la realización de investigaciones. ¿Cómo se selecciona a los participantes de un estudio? ¿Qué aspectos se deben tomar en cuenta cuando se decide cómo manipular una variable independiente? ¿Qué cosas deben preocuparnos cuando se mide una variable? ¿Qué se hace cuando el estudio ya está completo?

SELECCIÓN DE LOS PARTICIPANTES DE LA INVESTIGACIÓN

El centro de su investigación pueden ser niños, estudiantes universitarios, esquizofrénicos, ratas, pichones, conejos, primates e incluso cucarachas o lombrices; en cualquier caso, debe seleccionarse a los participantes de alguna forma. El método que se utilice para seleccionar a los participantes tiene implicaciones para la generalización de los resultados de la investigación.

Recuerde que en el capítulo 7 aprendimos que la mayoría de los proyectos de investigación requieren de la obtención de una muestra de participantes, a partir de una población de interés. La población se compone de todos los individuos en los que el investigador está interesado. Las muestras se pueden seleccionar de la población por medio de técnicas de muestreo probabilísticas o no probabilísticas. Cuando es importante describir a la población con precisión, se debe emplear el muestreo probabilístico. Por esto el muestreo probabilístico es tan importante cuando se realizan encuestas científicas. Sin embargo, gran parre de la investigación se interesa más en probar hipótesis sobre el comportamiento; la meta del estudio es la relación entre las variables estudiadas y la prueba de predicciones derivadas de teorías del comportamiento. En tales casos, se puede conseguir a los participantes de la forma más sencilla posible, utilizando métodos de muestreo accidentales o "por conveniencia", no probabilísticos. Se puede pedir a estudiantes de los cursos de introducción a la psicología que participen, que acudan a viviendas en su vecindario para encontrar personas que acepten resolver una prueba, o elegir una clase para someter a prueba a los niños, sencillamente porque *conoce* al maestro. No hay nada de malo con el uso de métodos como éstos, siempre y cuando reconozca que afectan la posibilidad de generalizar sus resultados a una población más grande. El tema de la generalización de resultados se estudia en el capítulo 14; a pesar de los problemas de generalización de resultados basados en muestras accidentales convenientes, existen muchas evidencias que sustentan la perspectiva de que *es posible generalizar* los hallazgos a otras poblaciones y situaciones.

También es necesario determinar el tamaño de la muestra. ¿Cuántos participantes necesitará para su estudio? En general, un incremento en el tamaño de la muestra aumenta las posibilidades de que sus resultados sean estadísticamente significativos, debido a que las muestras más grandes proporcionan estimativos más precisos de los valores poblacionales (véase la tabla 7.1). La mayoría de los investigadores se interesan en el tamaño de la muestra en el área que estudian, y seleccionan dimensiones de muestra que son típicas en dicha área.

En el capítulo 13 se describe un método más formal para la selección del tamaño de las muestras.

MANIPULACIÓN DE LA VARIABLE INDEPENDIENTE

Para manipular una variable independiente, usted debe construir una definición operacional de la misma (véase capítulo 4); es decir, debe convertir una variable conceptual en un conjunto de operaciones; instrucciones específicas, sucesos y estímulos que serán presentados a los participantes de la investigación. Además, las variables independiente y dependiente deben introducirse dentro del contexto del ambiente experimental total. A esto se le *conoce* como "preparación del terreno" (Aronson, Brewer y Carlsmith, 1985).

Preparación del terreno

Para preparar el terreno, generalmente se deben hacer dos cosas: dar a los participantes la información para el consentimiento informado requerido por el estudio, y explicarles por qué se realiza el experimento. En ocasiones se explican todos los fundamentos, aunque pocas veces se les comunica a los participantes las hipótesis reales. Por ejemplo, podría decir que va a realizar un experimento sobre memoria cuando, en realidad, va a estudiar un aspecto específico de la memoria (su variable independiente). Si los participantes conocen la meta del estudio, puede tratar de confirmar las hipótesis o comportarse de la manera más aceptable socialmente. Si usted considera que el engaño es necesario, tiene la obligación especial de comunicarlo a los participantes en el momento del informe, al concluir el experimento.

No existen reglas precisas para la preparación del terreno, con excepción de que el ambiente experimental debe parecer plausible a los participantes; tampoco hay reglas claras para traducir las variables conceptuales en operaciones específicas. La forma exacta de la manipulación de la variable depende de la propia variable y de los costos, así como de los aspectos prácticos y éticos de los procedimientos considerados.

Tipos de manipulación

Manipulación directa Por lo general, los investigadores pueden manipular una variable, con relativa sencillez, presentando material escrito, verbal o visual a los participantes. Este tipo de manipulación *directa* maneja las variables con instrucciones y presentación de estímulos. Éstos pueden presentarse de forma verbal, escrita, por medio de una cinta de video o una computadora. Veamos algunos ejemplos.

Labranche, Helweg-Larsen, Byrd y Choquette (1997) estudiaron el impacto de los folletos de promoción de la salud, cuando pidieron a mujeres que leyeran

un instructivo sobre el autoexamen mamario. En una condición, el folleto incluía únicamente texto; en otra se añadieron imágenes que describían el auto-examen. Las respuestas de las participantes a los dos folletos dependieron de qué tan cómodas se sintieron al leer los materiales con contenido sexual. Una de las preguntas interrogaba al respecto de si la mujer se creía capaz de realizar apropiadamente un autoexamen mamario. Las mujeres que se sentían incómodas ante material con contenido sexual se manifestaron más inseguras sobre su capacidad cuando leyeron el folleto con imágenes, que cuando leyeron el que contenía únicamente texto. El tipo de instructivo no afectó a las mujeres que se sentían cómodas ante material con contenido sexual.

Los estudios acerca de las decisiones de jurados suelen pedir a los participantes que lean una descripción de un juicio en la que se varía un elemento crucial de información. Bornstein (1998) estudió el efecto de la gravedad del daño sobre los juicios de responsabilidad de productos. Los participantes leyeron un caso en el que una mujer, que tomaba pastillas anticonceptivas, había sido diagnosticada con cáncer. En una condición de gravedad baja, el cáncer fue detectado de manera temprana, se extirpó un solo ovario, la mujer aún podía tener hijos y el pronóstico era bueno. En la condición de gravedad alta, el cáncer se detectó tardíamente, se extirparon ambos ovarios, de modo que era imposible un embarazo y el pronóstico era malo. En ambas condiciones se presentaron las mismas evidencias con respecto a la posible responsabilidad de las pastillas como causantes del cáncer, de modo que los juicios acerca de la responsabilidad del producto deberían ser los mismos en ambas condiciones. No obstante, la información sobre la gravedad afectó los juicios de responsabilidad: en la condición de alta gravedad, el 40 por ciento de los participantes consideró que el fabricante de las pastillas era responsable, en contraste con el 21 por ciento en la condición de poca gravedad.

La mayoría de las investigaciones sobre memoria se basan en manipulaciones directas. Por ejemplo, Coltheart y Langdon (1998) mostraron listas de palabras a un grupo de participantes y después midieron su recuerdo. Las listas de palabras eran fonológicamente diferentes: algunas incluían palabras en inglés con sonidos similares, como *cat, map* y *paz, y* otras contenían palabras distintas como *mop, pen y cow.* Los autores encontraron que los individuos recordaban con mayor precisión las listas con palabras distintas. En un estudio más complejo sobre la memoria, Reeve y Aggleton (1998) presentaron el libreto de un episodio futuro de una telenovela británica llamada *The Arcbers,* a un grupo de fanáticos ("expertos") y a otro grupo de personas que no estaban familiarizadas con el programa. En una condición, el libreto era típico de un episodio real: los Archer visitan un mercado de ganado. En la otra condición, el libreto era poco común: los Archer visitan una exposición de barcos. Los personajes y la estructura básica del programa fueron idénticos en ambas condiciones. Después de leer el libreto, se midió la retención que tenían los participantes sobre los detalles del episodio. Los investigadores encontraron que ser experto incrementaba la retención únicamente cuando se trataba de una historia típica. En la condición poco común, tanto los fanáticos como los individuos neófitos, tuvie-

ron niveles iguales de retención. Reeve y Aggleton concluyeron que los beneficios de ser un experto son muy limitados.

Como un último ejemplo de la manipulación directa, considere un estudio realizado por Petty, Cacioppo y Goldman (1981) sobre los efectos de la credibilidad del comunicador y el involucramiento personal en los cambios de actitudes. Los participantes eran estudiantes universitarios del último año, quienes leyeron sobre las razones de la aplicación obligatoria de un extenso examen para graduarse en la universidad. Para manipular la credibilidad, se les dijo que los argumentos habían sido escritos por un profesor de educación de Princeton University o por un estudiante de primer año de una universidad local. Los investigadores también manipularon el involucramiento personal cuando informaron a los estudiantes que se estaba considerando aplicar el examen ese mismo año (lo cual afectaría a los participantes del estudio) o diez años después. Los participantes en la condición de poco involucramiento mostraron mayor cambio de actitudes cuando el comunicador tenía una alta credibilidad, pero el crédito del comunicador no marcó una diferencia cuando los participantes estaban muy involucrados.

Descubrirá que la mayoría de las manipulaciones de las variables independientes, en muchas áreas de investigación, son de tipo directo. Los investigadores varían directamente la dificultad del material que debe aprenderse, los niveles de motivación, la forma en que se plantean las preguntas, las características de las personas a ser juzgadas y una gran variedad de otros factores.

Manipulación por etapas Otras manipulaciones son menos directas. En ocasiones, es necesario separar por etapas los sucesos que ocurren durante el experimento, para manipular la variable independiente de manera exitosa. Cuando esto sucede, se habla de una *manipulación por etapas o del suceso*.

La manipulación por etapas se utiliza con mayor frecuencia por dos razones. Primero, el investigador puede estar tratando de producir algún estado psicológico en los participantes, como frustración, enojo o una baja temporal de la autoestima; en segundo lugar, la manipulación por etapas puede ser necesaria para estimular alguna situación que ocurre en la vida real. Por ejemplo, Fazio. Cooper, Dayson y Johnson (1981) estudiaron el desempeño cognoscitivo bajo condiciones de demanda de tareas múltiples. Los participantes en una condición pasaron diez minutos haciendo la prueba de lectura de un manuscrito; los participantes en la otra condición realizaron la misma prueba, pero fueron interrumpidos por el experimentador en ciertos momentos, cuando les pedía que fueran a otra habitación a realizar otras tareas. Estas condiciones semejan ambientes de trabajo comunes de la vida real.

En la manipulación por etapas se suele utilizar un **cómplice** (en ocasiones llamado "confederado"). Generalmente, el cómplice parece ser otro participante en el experimento, pero en realidad forma parte de la manipulación (en el capítulo 3 hablamos de la utilidad de los cómplices). Un confederado puede ser útil para crear una situación social en particular. Por ejemplo, en un estudio sobre agresión, el cómplice y el participante se presentan al experimento, se les

dice que esperen en una habitación a que inicie el estudio. Durante el periodo de espera, el cómplice insulta al participante en una condición de "enojo", pero no lo hace en una condición "sin enojo". Entonces, el experimentador entra e informa a los dos individuos que interactuarán en una situación específica; la situación se diseña de tal modo que el participante real pueda aplicar un estímulo nocivo al cómplice, como un ruido fuerte o un olor desagradable. La cantidad de ruido u olor aplicados equivalen a la medida de la agresión; el investigador compara las cantidades aplicadas en la condición de enojo y en la condición sin enojo.

El experimento clásico de Asch (1956) sobre el conformismo proporciona otro ejemplo del uso de cómplices. Asch formó grupos de personas y les pidió que respondieran a una tarea en la que debían comparar líneas, como las que se muestran en la figura 9.1. ¿Cuál de las tres líneas es igual a la estándar? Aun cuando parece una tarea simple, Asch la volvió más interesante cuando hizo que los cómplices indicaran la misma respuesta incorrecta antes de preguntar al participante real; este procedimiento se repitió en varios ensayos, con distintos juicios de las líneas. Asch pudo demostrar lo fácil que es producir conformismo; los participantes mostraban conformidad con la mayoría en muchos de los ensayos, aun cuando la respuesta correcta era obvia. Finalmente, se pueden emplear cómplices en experimentos de campo, así como en investigación de laboratorio. Por ejemplo, Barón estudió la conducta de ayuda cuando hizo pasar por comprador a un cómplice en una tienda, el cual solicitaba cambio a compradores reales (Barón, 1997).

La manipulación por etapas demanda mucha ingenuidad e incluso dotes nistriónicas. Se emplea para involucrar a los participantes en una situación social que los individuos no perciben como experimento, sino como una experiencia real. Los investigadores suponen que el resultado será un comportamiento natural que refleje verdaderamente los sentimientos e intenciones de los participantes. Sin embargo, este tipo de procedimientos permite una gran cantidad de comunicación interpersonal sutil que es difícil poner en palabras; esto dificulta a otros investigadores replicar los experimentos. Además, una manipulación compleja es difícil de interpretar. Si durante el experimento sucedieron muchas cosas, *¿qué cosa* fue la responsable de los resultados? En general, es más fácil interpretar los resultados cuando la manipulación es relativamente directa. No obstante, la naturaleza de la variable que se estudia en ocasiones demanda procedimientos complicados.

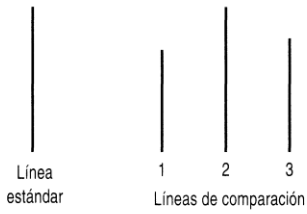


FIGURA 9.1
Ejemplo de una tarea del juicio de líneas de Asch.

Intensidad de la manipulación

El diseño experimental más sencillo incluye dos niveles de la variable independiente. Al planear el experimento, el investigador debe elegir estos niveles. Un principio general es hacer que la manipulación sea lo más intensa posible. Una manipulación intensa maximiza las diferencias entre los dos grupos e incrementa las posibilidades de que la variable independiente tenga un efecto, estadísticamente significativo, sobre la variable dependiente.

Como ejemplo, suponga que usted piensa que existe una relación lineal positiva entre la similitud de actitudes y la atracción ("Dios los hace y ellos se juntan"). Para realizar un experimento, usted podría arreglar las cosas de tal modo que los participantes se encuentren con otra persona, un cómplice. En un grupo, el cómplice y el participante comparten actitudes similares; en el otro grupo, el cómplice y el participante son distintos. Entonces, la similitud es la variable independiente y la atracción es la variable dependiente. Ahora debe tomar decisiones sobre la magnitud de la similitud. La figura 9.2 muestra la relación hipotética entre la similitud de las actitudes y la atracción, en diez niveles diferentes de similitud. El nivel 1 representa la menor cantidad de similitud, sin actitudes en común, y el 10 representa la mayor cantidad (todas las actitudes similares). Para lograr mayor intensidad en la manipulación, los participantes en un grupo se encuentran con un cómplice con un nivel de similitud de 1, y los del otro grupo con un cómplice con un nivel 10 de similitud. Esto podría dar como resultado la mayor diferencia en las medias de la atracción, una diferencia de 9 puntos. Una manipulación más débil, como utilizar los niveles 4 y 7, por ejemplo, resultaría en una diferencia menor entre medias.

Una manipulación intensa es especialmente importante en las primeras etapas de la investigación, cuando el científico está más interesado en demostrar si, de hecho, existe una relación. Si los primeros experimentos revelan una relación entre las variables, investigaciones posteriores pueden manipular, de manera

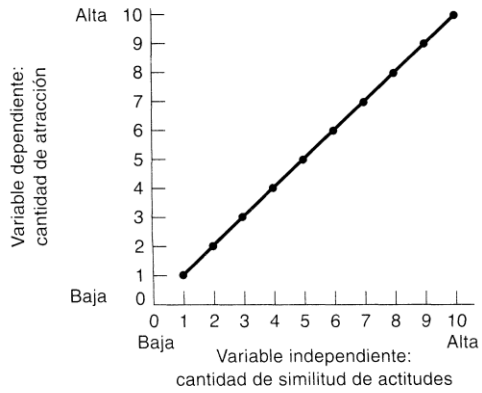


FIGURA 9.2
Relación entre la **similitud** de actitudes y la atracción.

sistemática, los otros niveles de la variable independiente para proporcionar una imagen más detallada de la relación.

El principio del uso de la manipulación más fuerte posible debe estar mediado por al menos dos aspectos. Primero, la manipulación más fuerte posible puede implicar una situación que ocurre raras veces, si acaso, en la vida real. Por ejemplo, una manipulación extremadamente fuerte del hacinamiento podría implicar colocar a tantas personas en una habitación, que ninguna de ellas pueda moverse, manipulación que podría afectar significativamente a una diversidad de conductas. Sin embargo, no podríamos saber si los resultados son similares a los que ocurren en situaciones más comunes, menos hacinadas, como en muchos salones de clases u oficinas.

El segundo aspecto es de tipo ético: la manipulación debe ser lo más fuerte posible, dentro de los límites de la ética. Una manipulación fuerte del temor o la ansiedad, por ejemplo, podría no ser posible debido al daño físico y psicológico potencial causado a los participantes.

Costo *de la manipulación*

El costo es otro factor en la decisión sobre la manipulación de la variable independiente. Los investigadores que tienen recursos económicos limitados tal vez no puedan adquirir equipo costoso, pagar el salario de los cómplices o pagar a los participantes en experimentos de larga duración. Asimismo, una manipulación en la que los individuos son tratados de manera individual requiere de mayor tiempo del experimentador, que una manipulación que permite tratar a muchos individuos en un mismo ambiente. Al respecto, una manipulación que utiliza la presentación directa de material escrito o verbal es menos costosa que una manipulación experimental compleja y por etapas. Algunas agen-cias gubernamentales y privadas otorgan fondos para investigación; debido a que muchas investigaciones son caras, es muy importante el apoyo público continuo.

MEDICIÓN DE LA VARIABLE DEPENDIENTE

En capítulos anteriores estudiamos diversos aspectos de la medición de variables, entre ellos la confiabilidad, la validez y la reactividad de las medidas, los métodos observacionales y la creación de medidas de autorreporte para cuestionarios y entrevistas. En este capítulo abordaremos aspectos de la medición que son especialmente relevantes para la investigación experimental.

Tipos de medidas

En la mayoría de los experimentos, existe al menos uno de tres tipos generales de variables dependientes: de autorreporte, conductuales o fisiológicas.

Medidas de autorreporte Los autorreportes pueden utilizarse para medir: actitudes, la atracción interpersonal, los juicios sobre las características de la personalidad de algún individuo, conductas deseables, estados emocionales, atribuciones acerca de por qué una persona se desempeñó bien o mal en una tarea, la confianza en los juicios propios, y muchos otros aspectos del pensamiento y la conducta humanos. Las más utilizadas son las escalas de puntuación con anclajes descriptivos (descriptores en los extremos). Por ejemplo, el estudio de Labranche y sus colaboradores (1997), descrito antes, pidió a las mujeres que respondieran en una escala de siete puntos, después de leer el folleto:

Siento que puedo realizar adecuadamente un autoexamen mamario.

Completamente en desacuerdo	_____	____	____	____	____	Completamente de acuerdo

Medidas conductuales Las medidas conductuales son observaciones directas del comportamiento. Como sucede con los autorreportes, es posible medir un número casi ilimitado de conductas. En ocasiones, el investigador puede registrar si ocurre o no cierta conducta; por ejemplo, si un individuo responde o no a una petición de ayuda, si comete un error en una prueba o elige involucrarse en una actividad y no en otra. Con frecuencia, el investigador debe decidir si registrará el número de ocasiones en que ocurre una conducta en un periodo dado —la *frecuencia* de una conducta—; la velocidad de la respuesta después de un estímulo —*el tiempo de reacción—;* o el tiempo que continuaba presentándose la conducta —una medida de *duración*—. La decisión de qué aspecto del comportamiento se medirá depende de cuál sea más relevante teóricamente para el estudio de un problema en particular, o de cuál medida surge, de manera lógica, de la manipulación de la variable independiente.

A veces, la naturaleza de la variable en estudio requiere un autorreporte o una medida conductual. La medida del comportamiento de ayuda es, casi por definición, una medida conductual, mientras que la medición de la percepción de las características de la personalidad de algún individuo requiere de un autorreporte. Sin embargo, para muchas variables tanto los autorreportes como las medidas conductuales podrían ser apropiadas. Así, la atracción podría medirse en una escala de puntuación o con una medida conductual de la distancia entre dos personas, o la cantidad de tiempo que pasan viéndose a los ojos. Cuando ambas opciones son posibles, se debe realizar una serie de estudios para indagar los efectos de una variable independiente sobre los dos tipos de medidas.

Medidas fisiológicas Las medidas fisiológicas son registros de respuestas corporales. Muchas de estas respuestas son observables; algunos ejemplos incluyen la **respuesta galvánica de la piel** (RGP), el **electromiograma** (EMG) y el **electroencefalograma** (EEG). La RGP es una medida de la activación emocional general y de la ansiedad; mide la conducción eléctrica de la piel, la cual cambia con la sudoración. El EMG mide la tensión muscular y con frecuencia se utiliza como una medida de tensión o estrés. El EEG mide la actividad eléctrica de las células

cerebrales; puede emplearse para registrar la activación cerebral general como respuesta a diferentes situaciones, la actividad en distintas partes del cerebro conforme ocurre el aprendizaje o la actividad cerebral durante distintas etapas del sueño.

La RGP, el EMG y el EEG se han utilizado desde hace mucho tiempo como indicadores fisiológicos de variables psicológicas importantes. Existen muchas otras medidas fisiológicas, entre ellas la temperatura, la frecuencia cardiaca y la información que puede obtenerse del análisis de sangre o de orina (véase Cacio-ppo y Tassinary, 1990). En años recientes, la imagen de resonancia magnética (IRM) se ha convertido en una herramienta cada vez más importante para los investigadores de la neurociencia del comportamiento. Una IRM proporciona una imagen de la estructura cerebral de un individuo. Permite a los científicos comparar la estructura cerebral de individuos con condiciones particulares (por ejemplo, un deterioro cognoscitivo, la esquizofrenia o el trastorno por hiperactividad y déficit de atención) con la estructura cerebral de individuos que no padecen la condición. Además, la llamada **IRM funcional** *(IRMf)* permite que los investigadores realicen búsquedas de áreas del cerebro, mientras el participante realiza una tarea física o cognoscitiva. Los datos brindan evidencias de los procesos cerebrales que están implicados en estas tareas. Por ejemplo, un investigador puede ver cuáles áreas del cerebro se encuentran más activas cuando el individuo realiza distintas tareas de memoria.

Sensibilidad de la variable dependiente

La variable dependiente debe ser lo suficientemente sensible como para detectar diferencias entre grupos. Una medida de la atracción interpersonal que pregunta "¿te gusta esta persona?", con una alternativa simple de respuesta "sí" o "no", es menos sensible que otra que cuestiona "¿cuánto te gusta esta persona?", en una escala de cinco o siete puntos. Con la primera medida, las personas tienden a ser agradables y a contestar afirmativamente, incluso si tienen sentimientos negativos hacia dicho individuo. La segunda medida permite una gradación de la atracción; una escala como éstas facilitaría la detección de diferencias en la cantidad de atracción.

El tema de la **sensibilidad** es particularmente importante cuando se mide el desempeño humano. La memoria puede medirse por medio del recuerdo, el reconocimiento o el tiempo de reacción; el desempeño en tareas cognoscitivas podría medirse examinando la velocidad o el número de errores durante una tarea de prueba de lectura; el desempeño físico puede medirse a través de varias tareas motoras que varían en su dificultad. En ocasiones, una tarea es tan fácil que cualquiera la hace bien, sin importar las condiciones manipuladas por la variable independiente. A este resultado se le conoce como **efecto de techo.** Aparentemente, la variable independiente no tiene un efecto sobre la medida dependiente, debido a que los participantes alcanzan rápidamente el máximo nivel de desempeño. El problema opuesto ocurre cuando una tarea es tan difícil que casi nadie puede realizarla bien, a esto se le llama **efecto de piso.**

La necesidad de considerar la sensibilidad de las medidas se ilustra muy bien en el estudio de Freedman y sus colaboradores (1971), sobre el hacinamiento, mencionado en el capítulo 4. El estudio examinó los efectos del hacinamiento con diversas medidas del desempeño en tareas cognoscitivas, y encontró que el hacinamiento no deteriora el desempeño. Usted podría concluir que el hacinamiento no tiene un efecto sobre el desempeño; no obstante, también es posible que las medidas fuesen demasiado fáciles o muy difíciles como para detectar un efecto del hacinamiento. De hecho, investigaciones posteriores mostraron que las tareas pudieron haber sido demasiado fáciles; cuando se pidió a los participantes que realizaran tareas más complejas, el hacinamiento dio como resultado un desempeño más bajo (Paulus, Anms, Seta, Schkade y Ma-tthews, 1976).

Medidas múltiples

Con frecuencia es deseable medir más de una variable dependiente. Una razón para utilizar medidas múltiples surge del hecho de que una variable puede medirse en una gran variedad de formas concretas (recuerde la explicación de las definiciones operacionales en el capítulo 4). Por ejemplo, en un estudio sobre las conductas relacionadas con la salud, los investigadores midieron el número de días de trabajo perdidos por problemas de salud, el número de visitas al médico y el uso de aspirina y tranquilizantes (Matteson e Ivancevich, 1983). También se podrían haber tomado medidas fisiológicas. Si la variable independiente tiene el mismo efecto sobre varias medidas de la misma variable dependiente, se incrementa nuestra confianza en los resultados. También es útil saber si la misma variable independiente afecta a algunas medidas, pero no a otras. Por ejemplo, una variable independiente diseñada para afectar la atracción podría tener un efecto sobre algunas medidas de la atracción (por ejemplo, el deseo de trabajar con una persona), pero no sobre otras (por ejemplo, el deseo de ser la pareja de esa persona). Los investigadores también pueden interesarse en estudiar los efectos de una variable independiente sobre varias conductas. Por ejemplo, un experimento sobre los efectos de una nueva técnica del manejo en el salón de clase podría examinar el desempeño académico, las tasas de interacción entre los compañeros de clase y la satisfacción con el maestro.

Cuando usted tiene más de una medida dependiente, surge la cuestión del *orden*. ¿Afectará el hecho de que algunas medidas se tomen primero? ¿Es posible que los resultados de una medida en particular difieran si se mide antes o después? Este aspecto es similar a los efectos del orden que se estudiaron en el capítulo 8, en el contexto de los diseños de medidas repetidas. Quizás el responder a las primeras medidas afecte de alguna forma a las respuestas de medidas posteriores; o tal vez se ponga más atención a las primeras medidas que a las últimas. Existen dos formas posibles de resolver este aspecto. Si parece que el problema es grave, el orden de la presentación de las medidas puede contrabalancearse por medio de las técnicas descritas en el capítulo 8. Es raro encontrar

evidencias de investigaciones previas en las que el orden haya constituido un problema grave. En este caso, la respuesta prudente es presentar primero las medidas más importantes y, después, las que no lo son tanto. Con este método, el orden no será un problema cuando se interpreten los resultados sobre las variables dependientes más relevantes. Incluso, cuando el orden puede llegar a constituir un problema potencial para algunas medidas, se minimiza la trascendencia general sobre el estudio.

La obtención de medidas múltiples en un solo experimento es valiosa cuando existen posibilidades de hacerlo. Sin embargo, tal vez sea necesario realizar una serie de experimentos para explorar los efectos de una variable independiente sobre diversas conductas.

Costo *de las medidas*

Otro aspecto que debe considerarse es el costo, ya que algunas medidas pueden ser más caras que otras. Las medidas de autorreporte, de papel y lápiz, suelen ser económicas; las medidas que requieren de observadores entrenados o de equipo elaborado pueden ser bastante costosas. Un investigador que estudia la conducta no verbal, por ejemplo, podría necesitar una cámara de video para registrar la conducta de cada participante en una situación. Después, dos o más observadores tendrían que ver las cintas para codificar conductas como el contacto visual, las sonrisas o el autocontacto. (Se requieren dos observadores para asegurarse de que las observaciones sean confiables.) Así, se tendrían que realizar gastos para el equipo y para el personal. Los aparatos de registro fisiológico también son costosos. Los investigadores requieren recursos de universidades o de agencias externas para llevar a cabo este tipo de investigación.

Ética

Siempre son importantes los aspectos éticos. Los investigadores deben ser extremadamente cuidadosos sobre posibles invasiones a la privacidad, y deben asegurarse siempre de garantizar todos los aspectos de la confidencialidad.

CONTROLES ADICIONALES

El diseño experimental básico incluye dos grupos: en el caso más sencillo, uno experimental que recibe la manipulación y un grupo control que no. El uso de un grupo control permite eliminar una variedad de explicaciones alternas basadas en la historia, la maduración, la regresión estadística, etc. En ocasiones se requieren procedimientos de control adicionales para regular otros tipos de explicaciones alternas. Dos aspectos generales de control se refieren a las expectativas que tienen los participantes y los experimentadores.

Control de las expectativas de los participantes

Características de la demanda Previamente señalamos que los experimentadores prefieren no informar a los participantes sobre las hipótesis específicas del estudio, o el propósito exacto de la investigación. Esto se debe al problema de las **características de la demanda** (Orne, 1962). Una característica de la demanda es cualquier aspecto de un experimento que pueda informar a los participantes los propósitos del estudio. La preocupación aquí es que cuando los participantes forman expectativas acerca de las hipótesis del estudio, hacen lo que sea necesario para confirmar las hipótesis; esto, por supuesto, suponiendo que están dispuestos a cooperar. Orne realizó investigaciones para demostrar que las personas son cooperadoras. Por ejemplo, pidió a un grupo de participantes que sumaran cifras en una hoja de papel; al finalizar, tomaban una tarjeta de un gran mazo para recibir más instrucciones. Cada tarjeta de instrucciones indicaba rasgar la hoja en 32 piezas y continuar a la siguiente página con números. ¡Los participantes continuaron esta ridicula tarea durante varias horas sin protestar ni preguntar! Aunque tal vez se pueda pensar en alguna situación en la que los individuos tenderían a cooperar poco, la propuesta de Orne sobre los participantes cooperadores parece ser generalmente correcta.

Una forma de controlar las características de la demanda es por medio del engaño; es decir, lograr que los participantes piensen que el experimento está estudiando una cosa, cuando en realidad está estudiando otra. El experimentador puede diseñar elaboradas historias falsas para explicar el propósito del estudio y disfrazar lo que en realidad se estudia; también podría disimular la medida dependiente utilizando una medida moderada, o colocándola entre un conjunto de **reactivos de relleno** no relacionados en un cuestionario. Otro método implica sencillamente evaluar si las características de la demanda constituyen un problema, cuando se pregunta a los participantes sus percepciones con respecto a los propósitos de la investigación. Es posible que los participantes no tengan una perspectiva precisa del propósito del estudio; o si algunos individuos adivinan las hipótesis, sus datos pueden analizarse por separado.

Las características de la demanda pueden eliminarse cuando las personas no están conscientes de que se está llevando a cabo un experimento, o de que su conducta está siendo observada. Así, los experimentos que se realizan en ambientes de campo y la investigación observacional en la que el observador está encubierto, o en la que se utilizan medidas moderadas, minimizan el problema de las características de la demanda.

Grupos placebo Un tipo especial de expectativa de los participantes surge en las investigaciones sobre los efectos de las drogas. Considere un experimento que investiga si un fármaco como el Prozac reduce la depresión. Un grupo de personas diagnosticadas con depresión recibe el fármaco y otro no. Ahora suponga que el grupo con el fármaco presenta una mejoría. No sabemos si ésta fue causada por las propiedades del fármaco o por las expectativas sobre sus efectos, lo que se denomina *efecto placebo.* En otras palabras, la simple administración de

una píldora o de una inyección puede ser suficiente para causar una mejoría en la conducta. Para controlar esta posibilidad se puede añadir un **grupo placebo.** Los participantes en él reciben una píldora o inyección que contiene una sustancia inerte e inocua; no reciben el fármaco que se administra a los miembros del grupo experimental. Si la mejoría es el resultado de las propiedades activas del fármaco, los participantes en el grupo experimental deben mostrar una mejoría mayor que los del grupo placebo. Si el grupo placebo mejora tanto como el experimental, entonces la mejoría se debe a un efecto placebo.

En ocasiones, las expectativas de los participantes son el centro de una investigación. Por ejemplo, Marlatt y Rohsenow (1980) hicieron un estudio para determinar cuáles efectos conductuales del alcohol se deben a la bebida, en oposición al impacto psicológico de saber que se está ingiriendo alcohol. El diseño experimental para examinar estos efectos incluyó cuatro grupos: 1) sin expectativa de alcohol —sin consumo de alcohol—; 2) sin expectativa de alcohol —con consumo de alcohol—; 3) con expectativa de alcohol —sin consumo de alcohol— y 4) con expectativa de alcohol —con consumo de alcohol—. A esto se le denomina *diseño placebo balanceado.* La investigación de Marlatt y Rohsenow sugiere que la creencia de haber consumido alcohol es un determinante con mayor peso de la conducta, que el alcohol mismo. Es decir, las personas que creyeron haber consumido alcohol (grupos 3 y 4) se comportaron de manera muy similar, a pesar de que a los participantes del grupo 3 no se les dio alcohol.

En ciertas áreas de investigación el uso de grupos placebo de control tiene implicaciones éticas. Suponga que se encuentra estudiando un tratamiento que tiene un efecto positivo para la gente (por ejemplo, la reducción de migrañas o el alivio de síntomas depresivos). Es importante emplear procedimientos experimentales cuidadosos para asegurarse de que el tratamiento tiene resultados y que se eliminen las explicaciones alternas del efecto, incluso un efecto placebo. Sin embargo, también es importante ayudar a las personas que se encuentran en el grupo control. Estos participantes deben recibir tratamiento tan pronto como hayan completado su parte del estudio.

Los efectos placebo son reales y deben estudiarse con seriedad en muchas áreas de investigación. En la actualidad existe una numerosa cantidad de estudios y una gran polémica sobre la magnitud en que los efectos benéficos de medicamentos antidepresivos, tales como el Prozac, se deban a un efecto placebo (por ejemplo, Kirsch, Moore, Scoboria y Nicholls, 2002).

Control de las expectativas del investigador

Generalmente, los investigadores conocen los propósitos del estudio y, por lo tanto, pueden desarrollar expectativas sobre la forma en que los participantes deben responder. Estas expectativas pueden, a su vez, sesgar los resultados. A este problema general se le llama **sesgo del experimentador** o **efectos de las expectativas** (Rosenthal, 1966, 1967, 1969).

Los efectos de las expectativas pueden ocurrir cuando el experimentador conoce en qué condición participa cada individuo. Existen dos fuentes poten-

cíales del sesgo del experimentador. Primero, el investigador puede tratar, sin darse cuenta, de manera distinta a los participantes en las diversas condiciones del estudio. Por ejemplo, podría poner mayor énfasis a ciertas palabras cuando lee las instrucciones a uno de los grupos, pero no con el otro; o podría sonreír más al interactuar con los individuos de una de las condiciones. La segunda fuente de sesgo puede presentarse cuando los experimentadores registran las conductas de los participantes; puede haber diferencias sutiles en la manera en que el investigador interpreta y registra las conductas.

Investigación sobre los efectos de las expectativas Los efectos de las expectativas se han estudiado de diversas maneras. Quizás la primera demostración del problema sea el caso de Hans *el listo,* cuya, supuesta inteligencia era pura ilusión, según lo demostró Pfungst (1911). Robert Rosenthal describe a Hans *el listo:*

> Se recordará a Hans como el caballo listo que podía resolver problemas matemáticos y de armonías musicales con habilidad y gracia, con el simple hecho de golpear el piso con su pezuña. Un grupo de expertos atestiguó que Hans, cuyo propietario no obtuvo beneficios económicos de su talento, no recibía claves de quienes lo cuestionaban. Por supuesto, Pfungst demostró después que esto no era así, que Hans iniciaba y finalizaba los golpes de acuerdo a señales de pequeños movimientos de la cabeza y de los ojos. Cuando se le planteaba una pregunta, quien la formulaba siempre observaba su pezuña, situación muy natural, ya que era la forma para determinar si la respuesta de Hans era correcta. Entonces, se descubrió que cuando el caballo se aproximaba al número correcto de golpes, el indagador movía inadvertidamente su cabeza o sus ojos hacia arriba, lo suficiente como para que Hans pudiera discriminar la señal, aunque no lo suficiente para que incluso observadores de animales entrenados o psicólogos pudieran verlo.[1]

Si un caballo listo puede responder a señales sutiles, es razonable suponer que humanos inteligentes también puedan hacerlo. De hecho, las investigaciones han demostrado que las expectativas del experimentador pueden comunicarse a los seres humanos por medios verbales y no verbales (Duncan, Rosenberg y Finklestein, 1969; Jones y Cooper, 1971). Un ejemplo de una investigación más sistemática sobre los efectos de las expectativas es el estudio de Rosenthal (1966). En este experimento, estudiantes graduados entrenaron ratas que les fueron descritas como provenientes de cepas genéticamente "brillantes para los laberintos" o "ineptas para los laberintos". En realidad, los animales provenían de la misma cepa y habían sido asignados aleatoriamente a las categorías de brillantes e ineptos; sin embargo, las ratas "brillantes" *si* se desempeñaron mejor que las ratas "ineptas". Los resultados deben haber sido causados por diferen-

[1] Rosenthal, R. (1967). Convert communication in the psychological experiment. *Psy-chological Bulletin, 67,* pp. 356-367. Derechos reservados, 1967, American Psychological Asso-ciation. Reproducido bajo permiso.

cias sutiles en la forma en que los estudiantes trataron a las ratas o registraron su conducta. La generalización de este hallazgo, en particular, se *conoce* como "expectativa del maestro". Las investigaciones han demostrado que decirle a un maestro que un alumno florecerá intelectualmente durante el año siguiente, da como resultado un incremento en el puntaje de CI del alumno (Rosenthal y Jacobson, 1968). En resumen, las expectativas de los maestros pueden influir en el desempeño de los estudiantes.

El problema de la influencia de las expectativas sobre la evaluación de la conducta, se ilustra muy bien en un experimento realizado por Langer y Abelson (1974). A un grupo de psicólogos clínicos le mostró una cinta de video con una entrevista, en la que la persona entrevistada fue descrita como un solicitante de empleo, o como un paciente; en realidad, todos vieron la misma grabación. Los psicólogos calificaron a la persona como más "trastornada" cuando pensaron que se trataba de un paciente, que cuando fue descrita como solicitante de empleo.

Soluciones al problema de las expectativas Claramente, las expectativas del experimentador pueden influir en los resultados de las investigaciones. ¿Cómo puede resolverse este problema? Por fortuna, hay diversas formas de minimizar los efectos de las expectativas. En primer lugar, los experimentadores deben estar bien entrenados y deben practicar un comportamiento consistente con todos los participantes. Los beneficios del entrenamiento se demostraron en el estudio de Langer y Abelson con los psicólogos clínicos. El sesgo de considerar al "paciente" como un individuo trastornado fue mucho menor entre los terapeutas con orientación conductual, que entre los tradicionales. Tal vez el entrenamiento de los terapeutas conductuales los condujo a enfocarse más en la conducta real de la persona, de modo que se vieron menos influidos por las expectativas que provenían de la etiqueta de "paciente".

Otra solución es llevar a cabo todas las condiciones del experimento de manera simultánea, con el fin de que la conducta del experimentador sea la misma con todos los participantes. No obstante, esta solución sólo es posible bajo ciertas circunstancias, como cuando el estudio puede realizarse con el uso de materiales impresos o cuando las instrucciones que el experimentador da a los participantes son iguales para todos.

Los efectos de las expectativas también se minimizan cuando se automatizan los procedimientos. Como se señaló antes, es posible manipular variables independientes y registrar respuestas por medio de computadoras; con procedimientos automatizados es poco probable que las expectativas del experimentador afecten los resultados.

Una solución final es emplear experimentadores que desconozcan las hipótesis investigadas. En estos casos, la persona que lleva a cabo el estudio o realiza las observaciones está ciega con respecto al tema de estudio o al grupo al que pertenece el participante. Este procedimiento se originó en las investigaciones con fármacos, por medio del uso de grupos placebo. En un experimento de *ciego simple,* el participante no sabe si se le está administrando un placebo o un fárma-

co verdadero; en un experimento de *doble ciego,* ni el participante ni el experimentador conocen qué tratamiento se está aplicando. Para utilizar un procedimiento en el que el experimentador u observador desconozca las hipótesis o el grupo al que pertenecen los participantes, se debe contratar a otras personas para que lleven a cabo el experimento y realicen las observaciones.

Puesto que los investigadores están conscientes del problema de los efectos de las expectativas, generalmente se incorporan soluciones como las que acabamos de describir a los procedimientos de estudio. Si un estudio tiene un problema potencial de efectos de las expectativas, los investigadores casi con seguridad lo notarán e intentarán replicar el experimento con procedimientos que los controlen. Los procedimientos empleados en investigaciones científicas deben definirse con precisión, de manera que otros experimentadores puedan replicarlos; esto permite que otros científicos profundicen en investigaciones previas. Además, es un mecanismo autocorrectivo que asegura que se detecten las fallas metodológicas. En el capítulo 14 se estudiará la importancia de la replicación.

CONSIDERACIONES ADICIONALES

Hasta ahora hemos tratado varios de los factores que un investigador toma en cuenta cuando planea un estudio. La realización del estudio y el análisis de los resultados son procesos que requieren tiempo. Antes de iniciar la investigación, el científico debe asegurarse lo más posible de que todo se hará correctamente. Una vez que se ha diseñado el estudio, existen algunos procedimientos adicionales que pueden mejorarlo.

Propuestas de investigación

Después de dedicar un gran esfuerzo a la planeación del estudio, el investigador redacta una propuesta de investigación. Ésta incluye una revisión de la literatura que da un fundamento al estudio. El objetivo es explicar con claridad las razones de la investigación; para responder qué preguntas está diseñando el estudio. Se incluyen los detalles de los procedimientos que se emplearán para probar la idea, así como los planes para el análisis de los datos. Una propuesta de investigación es muy similar a las secciones de introducción y método de un artículo científico.

Este tipo de propuestas deben incluirse en las solicitudes para obtener fondos de investigación; los comités de revisión ética también requieren algún tipo de propuesta (véase el capítulo 3 para mayor información sobre los consejos institucionales de revisión). Cuando se planea cualquier proyecto de investigación, es bueno preparar una propuesta. El simple hecho de poner sus pensamientos en el papel le ayuda a organizar y sistematizar las ideas. Además, usted puede mostrar la propuesta a sus amigos, colegas, profesores y otras personas interesadas que puedan brindar una retroalimentación útil sobre sus procedimientos.

Es probable que ellos vean problemas que usted no ha observado o pueden darle ideas para mejorar el estudio.

Estudios piloto

Cuando el investigador finalmente ha decidido todos los aspectos específicos del procedimiento, puede realizar un **estudio piloto,** en el que se lleva cabo un "ensayo" con un pequeño número de participantes. El estudio piloto revelará si los individuos comprenden las instrucciones, si es factible el ambiente experimental total, si se plantean algunas preguntas confusas, etcétera.

En algunos estudios piloto, al finalizar el experimento, se pide a los participantes que ofrezcan detalles sobre su experiencia. Otro método es utilizar el protocolo del "pensamiento en voz alta" (descrito en el capítulo 7), en el cual se instruye a los participantes del estudio piloto para que verbalicen sus pensamientos sobre todo lo que está sucediendo durante el experimento. Este tipo de recursos le dan al investigador la oportunidad de hacer cualquier cambio necesario en el procedimiento, antes de llevar a cabo el estudio completo. Además, un estudio piloto permite que los experimentadores que recolectan los datos se sientan cómodos con su papel y que estandaricen sus procedimientos.

Verificación de la manipulación

La verificación de la manipulación es un intento de evaluar directamente si la manipulación de la variable independiente tiene el efecto buscado en los participantes. Las verificaciones de la manipulación ofrecen evidencias de la validez de constructo de la manipulación (la validez de constructo se estudió en el capítulo 4). Si usted está manipulando la ansiedad, por ejemplo, una verificación de la manipulación indicaría si los participantes del grupo de alta ansiedad en realidad estaban más ansiosos que los individuos en la condición de baja ansiedad. La verificación de la manipulación puede incluir un autorreporte de ansiedad, una medida conductual (como el número de movimientos de los brazos y de las manos) o una medida fisiológica. Por lo tanto, todas las verificaciones de la manipulación revisan si la manipulación de la variable independiente fue en efecto una operacionalización exitosa de la variable conceptual del estudio. Considere, por ejemplo, la manipulación del atractivo físico como variable independiente. En un experimento, los participantes responden a una persona que se supone es percibida como atractiva o poco atractiva. La corroboración de la manipulación, en este caso, determinaría si los participantes califican a la persona con alto atractivo como tal.

Las verificaciones de la manipulación son particularmente útiles en el estudio piloto, con el fin de decidir si la manipulación de la variable independiente está teniendo el efecto buscado. Si la variable independiente no es eficaz, se pueden cambiar los procedimientos. Sin embargo, siempre es importante realizar una verificación en el experimento real. Puesto que esto podría distraer a los

participantes o informarles acerca del propósito del experimento, generalmente es mejor administrar la medida de la verificación de la manipulación casi al final del experimento; en la mayoría de los casos, esto se hace después de medir las variables dependientes y antes de la sesión de informe.

La revisión de la manipulación tiene dos ventajas. En primer lugar, si la verificación muestra que la manipulación no fue eficaz, usted se habrá ahorrado el gasto de realizar el experimento real y puede centrar su atención en cambiar la manipulación para hacerla más eficiente. Por ejemplo, si la verificación índica que ni el grupo de alta ansiedad ni el de baja ansiedad estaban realmente ansiosos, usted podría cambiar los procedimientos para incrementar la ansiedad en la condición de alta ansiedad.

En segundo lugar, la verificación es ventajosa si no obtiene resultados significativos; es decir, si los resultados indican que no existe una relación entre las variables independiente y dependiente. La verificación puede identificar si los resultados no significativos se deben a un problema en la manipulación de la variable independiente. Si su manipulación no es exitosa, lo más razonable es que no tendrá resultados significativos. Si ambos grupos tienen los mismos niveles de ansiedad, después de manipular esta variable, la ansiedad no puede tener efecto alguno sobre la medida dependiente. ¿Qué sucede si la verificación indica que la manipulación fue exitosa, pero usted obtiene resultados no significativos? Entonces usted sabe, al menos, que los resultados no se debieron a un problema con la manipulación; las razones de no encontrar una relación radican en alguna otra cosa. Quizá su medida dependiente es inadecuada, o tal vez no existe en realidad una relación entre las variables.

Informe

La importancia del informe se comentó en el capítulo 3, en el contexto de las consideraciones éticas. Una vez que se han reunido todos los datos, es común que se lleve a cabo una sesión de informe, que representa una oportunidad para que el investigador interactúe con los participantes y discuta las implicaciones éticas y educativas del estudio.

La sesión de informe también puede brindar una oportunidad para aprender más acerca de lo que los participantes pensaron durante el experimento. Los investigadores pueden preguntarles cuál creían que era el propósito del experimento, cómo interpretaron la manipulación de la variable independiente y lo que pensaron cuando respondieron a las medidas dependientes. Esta información puede ser útil para interpretar los resultados y planear futuros estudios.

Finalmente, los investigadores pueden pedir a los participantes que se abstengan de hablar sobre el estudio con otras personas. Este tipo de solicitudes suelen hacerse cuando habrá más participantes, y cuando pueden comunicarse con otros individuos en las clases o en los dormitorios. Las personas que ya participaron están conscientes de los propósitos y procedimientos generales; es importante que estos individuos no provoquen expectativas sobre el estudio a otros posibles participantes.

ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Una vez que se han reunido los datos, el siguiente paso es analizarlos. Los análisis estadísticos de los datos se realizan para permitir que el investigador examine e interprete el patrón de los resultados obtenidos en el estudio. El análisis estadístico ayuda a que decida si en realidad existe una relación entre las variables independiente y dependiente; la lógica que subyace al uso de las pruebas estadísticas se estudia en el capítulo 13. No es el propósito de este libro enseñar métodos estadísticos; no obstante, en el apéndice B se incluyen los cálculos de varias pruebas estadísticas.

COMUNICAR A OTROS LA INVESTIGACIÓN

El paso final es redactar un reporte que detalle las razones de la investigación, la forma en que se reclutó a los participantes, los procedimientos utilizados y los hallazgos. En el apéndice A se muestra una descripción de la redacción de este tipo de reportes. Después de haber escrito el reporte, ¿qué se hace con él? ¿De qué manera se comunican los hallazgos a otras personas? Lo más común es que los descubrimientos de las investigaciones se presenten como artículos científicos o documentos en congresos científicos. En cualquier caso, dos o más revisores reconocidos evalúan el documento y deciden si se acepta para ser publicado o para ser presentado en el congreso.

Congresos profesionales

Los congresos organizados por asociaciones profesionales constituyen oportunidades importantes para que los investigadores presenten sus hallazgos a otros científicos y al público en general. Las asociaciones profesionales nacionales y regionales, como la Asociación Psicológica Americana (American Psychological Association, APA) y la Sociedad Psicológica Americana (American Psychological Society, APS) organizan y realizan congresos anuales en los que psicólogos *y* estudiantes de psicología presentan sus propias investigaciones y conocen los estudios más recientes de sus colegas. Algunas veces se hacen presentaciones verbales ante una audiencia. Sin embargo, las sesiones de cartel son más comunes; aquí, los investigadores muestran carteles que resumen la investigación y están disponibles para comentar su estudio con los demás.

Artículos científicos

Como se señaló en el capítulo 2, existen muchas revistas en las que se publican artículos científicos. No obstante, la cantidad de revistas es reducida, en comparación con la cantidad de reportes escritos; por lo tanto, no es fácil publicar una investigación. Cuando un investigador envía un artículo a una revista, dos o

más revisores lo leen y recomiendan su aceptación (con las estipulaciones que deben seguir las revisiones) o su rechazo. A este proceso se le llama *revisión de colegas, o depares,* y es muy importante para asegurarse de que las investigaciones sean sometidas a una revisión externa cuidadosa antes de su publicación. Aproximadamente el 90 por ciento de los artículos que se envían a las revistas más prestigiosas son rechazados. Muchos de estos artículos se remiten a otras revistas y son aceptados, eventualmente, para su publicación, aunque muchos estudios nunca lo logran. Esto no es necesariamente malo, simplemente significa que el proceso de selección separa las investigaciones de alta calidad de aquéllas de menor calidad.

En este capítulo se describieron muchas de las decisiones que deben tomarse cuando se planea un experimento. Las explicaciones se enfocaron en los experimentos que utilizan el diseño experimental más sencillo, con una sola variable independiente. En el siguiente capítulo se describen diseños experimentales más complejos.

Términos de estudio

Características de la demanda
Cómplice
Efecto de piso
Efecto de techo
Efectos de las expectativas (sesgo del experimentador)
Electroencefalograma Electromiograma
Estudio piloto Grupo placebo Informe
Intensidad de la manipulación IRM
IRM funcional Manipulación directa Manipulación por etapas
Medida conductual Medida de autorreporte Medida fisiológica
Procedimiento de doble ciego
Reactivos de relleno Respuesta galvánica de la piel Sensibilidad
Procedimiento de ciego simple
Verificación de la manipulación

Preguntas de repaso

1. ¿Cuál es la diferencia entre la manipulación directa y la manipulación por etapas, de una variable independiente?
2. Indique las diferencias entre los tipos generales de variables dependientes.
3. ¿Qué es la sensibilidad de una medida? ¿Cuáles son los efectos de techo y de piso?
4. Comente las formas en que pueden utilizarse las computadoras para realizar un experimento.
5. ¿Qué son las características de la demanda? Describa formas que se utilizan para minimizar las características de la demanda.

6. ¿Cuál es la razón del uso de un grupo placebo?
7. ¿Qué son los efectos de las expectativas del experimentador? Indique algunas soluciones al problema del sesgo del experimentador.
8. ¿Qué métodos pueden emplearse para depurar un experimento?
9. ¿Qué es un estudio piloto?
10. ¿Qué es la verificación de la manipulación? ¿De qué manera ayuda al investigador a interpretar los resultados de un experimento?
11. ¿Qué hace el investigador con los hallazgos después de haber completado un proyecto de investigación?

Actividades

1. El doctor Turk estudió la relación entre la edad y la comprensión de la lectura, y específicamente predijo que las personas más grandes mostrarían menor comprensión que las más jóvenes. Turk estaba particularmente interesado en la comprensión del material que está disponible en la prensa en general. Grupos de participantes de 20, 30, 40 y 50 años de edad leyeron un capítulo de un libro del físico Stephen W. Hawking (1988), llamado *A Brief History of Time: From the BigBang to the Black Holes* (el cual estaba en la lista de los libros más vendidos en esa época). Después de leer el capítulo, se aplicó una medida de comprensión a los participantes. Los resultados indicaron que no había una relación entre la edad y las puntuaciones de comprensión; todos los grupos de edad tuvieron puntuaciones igualmente bajas en la medida de comprensión. ¿Por qué cree que no se encontró relación alguna? Identifique, al menos, dos posibles razones.
2. Recuerde el experimento sobre la facilitación de la comunicación de niños con autismo, descrito en el capítulo 2 (Montee, Miltenberger y Wittrock, 1995). Interprete los hallazgos de dicho estudio en términos de los efectos de las expectativas del experimentador.
3. Su grupo de laboratorio ha sido asignado para diseñar un experimento que investigue el efecto del tiempo de estudio sobre una tarea de recuerdo. Hasta ahora, su grupo ha organizado el siguiente plan: "los participantes se asignarán aleatoriamente a dos grupos. Los individuos en un grupo estudiarán una lista de cinco palabras durante cinco minutos, y los del otro grupo estudiarán la misma lista durante siete minutos. Inmediatamente después de estudiar, los participantes leerán una lista de diez palabras y encerrarán en un círculo aquellas que aparecían en la lista del estudio original". Mejore este experimento, dando razones específicas para cualquier cambio.
4. Si usted estuviese investigando variables que afectan la conducta de ayuda, ¿utilizaría una manipulación directa o una por etapas? ¿Por qué?

5. Diseñe un experimento que utilice la manipulación por etapas para probar la hipótesis de que cuando las personas están de buen humor son más propensas a contribuir con causas caritativas. Incluya una verificación de la manipulación en su diseño.
6. En un estudio piloto, la doctora Mori llevó a cabo una verificación de la manipulación y no encontró diferencias significativas entre las condiciones experimentales. ¿Debe ella continuar con el experimento? ¿Qué debe hacer a continuación? Explique sus recomendaciones a la doctora Mori.
7. Redacte un documento de informe que leería a los participantes del estudio de Asch.

10

Diseños experimentales complejos

Incremento del número de niveles de una variable independiente

Incremento del número de variables independientes: diseños factoriales

Interpretación de los diseños factoriales

Efectos principales

Interacciones Diseños factoriales con manipulación y sin manipulación de variables Interacciones y variables moderadoras Resultados de un diseño factorial 2x2 Interacciones y efectos principales simples

Efecto principal simple del tipo de entrevistador

Efecto principal simple del tipo de pregunta

Procedimientos de asignación y diseños factoriales

Grupos independientes

Medidas repetidas

Diseño factorial mixto con el uso de asignación combinada Incremento del número de niveles de una variable independiente Incremento del número de variables independientes en un diseño factorial

Términos de estudio

Preguntas de repaso

Actividades

Hasta ahora nos hemos enfocado principalmente en el diseño experimental más simple, en el que se manipula una variable independiente con dos niveles y se mide una variable dependiente. Este sencillo diseño nos permite examinar aspectos importantes de una investigación, como la validez interna y los procedimientos de asignación de los participantes a las diferentes condiciones. Sin embargo, con frecuencia los científicos investigan problemas que demandan diseños más complejos. Este tipo de diseños constituye el tema de este capítulo.

INCREMENTO DEL NUMERO DE NIVELES DE UNA VARIABLE INDEPENDIENTE

En el diseño experimental más sencillo existen únicamente dos niveles de la variable independiente. No obstante, un investigador puede verse ante la necesidad de diseñar un experimento con tres o más niveles, por varias razones. Primero, un diseño que incluye sólo dos niveles de la variable independiente no puede proporcionar mucha información sobre la relación exacta que existe entre las variables independiente y dependiente. Por ejemplo, la figura 10.1 muestra los resultados de un experimento hipotético sobre la relación entre la motivación y el desempeño en una tarea motora. La línea sólida describe los resultados cuando existen únicamente dos niveles: no hay recompensa por un desempeño satisfactorio y la promesa de $4.00 por un alto desempeño. Puesto que solamente hay dos niveles, la relación puede describirse únicamente por medio de una línea recta. No sabemos cómo sería la relación si se incluyeran otras cantidades en los niveles de la variable independiente. La línea punteada de la figura 10.1 indica los resultados cuando se prometen $1.00, $2.00 y $3.00. Este resultado representa una descripción más precisa de la relación que existe entre la cantidad de la recompensa prometida y el desempeño. En el experimento hipo-

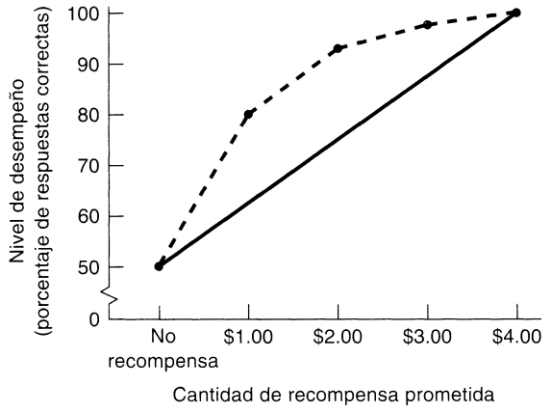


FIGURA 10.1
Resultados de un experimento hipotético: función lineal *vs.* función monotóníca positiva.

tético, la cantidad de recompensa es muy eficaz para incrementar el desempeño hasta un punto, después del cual únicamente ocurren mejorías modestas en el desempeño, acompañadas de aumentos en la recompensa. Por lo tanto, se trata de una relación monotónica positiva, y no de una relación estrictamente lineal (véase el capítulo 4). Un experimento con sólo dos niveles no puede producir información tan precisa.

En el capítulo 4 explicamos que las variables en ocasiones se relacionan de manera curvilínea o no monotónica, es decir, que cambia la dirección de la relación. La figura 10.2 presenta un ejemplo de una relación curvilínea; a esta forma particular de relación curvilínea se le denomina *U invertida*. Un diseño experimental que sólo posee una variable independiente con dos niveles, no puede detectar relaciones curvilíneas entre variables. Si estamos prediciendo una relación curvilínea, se deben utilizar tres niveles por lo menos. Como la figura 10.2 índica, si sólo se hubiesen utilizado los niveles 1 y 3 de la variable independiente, no se habría detectado ninguna relación entre las variables. En psicología existen muchas relaciones curvilíneas de este tipo. La relación que existe entre la aparición del temor y el cambio de actitudes es un ejemplo. El aumento de la cantidad de temor, propiciado por un mensaje persuasivo, cambia las actitudes hasta un nivel moderado de temor; mayores incrementos en el temor reducen el cambio de actitudes.

Finalmente, los investigadores suelen interesarse en comparar más de dos grupos. Suponga que usted desea saber si jugar con un animal tiene efectos benéficos para residentes de asilos. Podría establecer dos condiciones, como un grupo control sin animales y un grupo que puede jugar con un perro diariamente. Sin embargo, es probable que usted se interese también en conocer los efectos de un gato y un pájaro, por lo cual podría añadir estos dos grupos a su estudio. Asimismo, podría interesarse en comparar los efectos de un perro grande y uno chico, además de la condición control sin animales. En un estudio real sobre técnicas de reducción del estrés, realizado por Bruning y Frew (1987), se asignaron aleatoriamente empleados a uno de cuatro grupos: ejercicio, entrenamiento de habilidades de manejo de estrés, meditación y un grupo control sin tratamiento. En este estudio, las tres técnicas produjeron decrementos en la presión sanguínea y en el pulso.

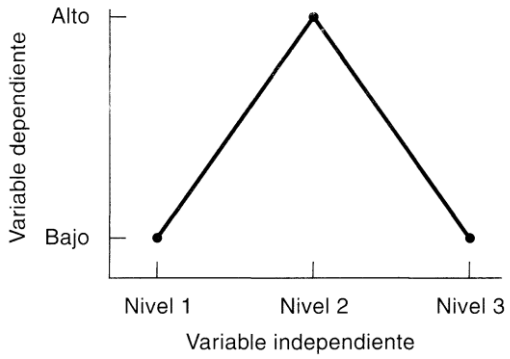


FIGURA 10.2

Relación curvilínea de U invertida. *Nota:* se requieren al menos tres niveles de la variable independiente para que se presenten relaciones curvilíneas.

INCREMENTO DEL NÚMERO DE VARIABLES INDEPENDIENTES: DISEÑOS FACTORIALES

Con frecuencia, los investigadores manipulan más de una variable independiente en un solo experimento. Generalmente, operan dos o tres variables independientes de manera simultánea. Este tipo de diseño experimental es una aproximación a las situaciones reales, en las que las variables independientes no existen por sí mismas. Los investigadores reconocen que en cualquier situación operan muchas variables que afectan el comportamiento, por lo cual diseñan experimentos que incluyen más de una variable independiente.

Los **diseños factoriales** incluyen más de una variable independiente *(o factor).* En un diseño factorial todos los niveles de cada variable independiente se combinan con todos los niveles de las otras variables independientes. El diseño factorial más simple, *conocido* como diseño factorial 2x2 (dos por dos), incluye dos variables independientes, cada una de ellas con dos niveles.

Un experimento de Smith y Ellsworth (1987) ilustra el diseño factorial 2x2. Estos autores estudiaron los efectos de plantear preguntas sesgadas sobre la precisión del relato de testigos. Los participantes en el experimento primero observaron la filmación de un robo; luego, los experimentadores les plantearon preguntas acerca de lo que vieron. Una variable independiente era el tipo de pregunta: sesgada o sin sesgo. La segunda variable independiente era el conocimiento que el entrevistador tenía sobre el crimen: la persona que preguntaba había visto la filmación una sola vez (un entrevistador "inexperto") o la había visto varias veces (un entrevistador "experto").

Este diseño 2 x 2 se deriva en cuatro condiciones experimentales: 1) entrevistador experto-preguntas sesgadas; 2) entrevistador experto-preguntas sin sesgo; 3) entrevistador inexperto-preguntas sesgadas y 4) entrevistador inexper-to-preguntas sin sesgo. Un diseño 2 x 2 siempre incluye cuatro grupos. El formato general para describir diseños factoriales es

Número de niveles de la primera VI	X	Número de niveles de la segunda VI	X	Número de niveles de la tercera VI

y así sucesivamente. Un diseño con dos variables independientes, una con dos niveles y la otra con tres niveles dan como resultado un diseño factorial 2x3, por lo que existen seis condiciones en el experimento. Un diseño 3x3 incluye nueve condiciones.

Interpretación de los diseños factoriales

Los diseños factoriales producen dos tipos de información. El primero es la información sobre el efecto de cada variable independiente de manera aislada: el **efecto principal** de una variable independiente. En un diseño que incluye dos variables independientes existen dos efectos principales, uno de cada variable

independiente. El segundo tipo de información se denomina **interacción.** Si existe interacción entre dos variables independientes, el efecto de una de ellas depende del nivel particular de la otra. En otras palabras, el efecto que una variable independiente tiene sobre la variable dependiente depende del nivel de la otra variable independiente. Las interacciones son una nueva fuente de información que no puede obtenerse en un diseño experimental simple, en el que sólo se manipula una variable independiente.

Para ilustrar los efectos principales y las interacciones, observemos los resultados del estudio de Smith y Ellsworth sobre la precisión del relato de testigos. La tabla 10.1 incluye un método común para la presentación de resultados de los diversos grupos en un diseño factorial. La cifra en cada celda representa el porcentaje medio de errores cometidos en las cuatro condiciones.

TABLA 10.1 Diseño factorial 2X2: resultados del experimento del relato de testigos.

Tipo de entrevistador (variable independiente *B)*	Tipo de pregunta (variable independiente *A)*		Medias totales (efecto principal de *B)*
	Sin sesgo	Sesgada	
Experto	13	41	27.0
Inexperto	13	18	15.5
Medias totales (efecto principal de *A)*	13.0	29.5	

Efectos principales Un efecto principal es aquel que cada variable tiene por sí misma. El efecto principal de la variable independiente A, el tipo de pregunta, es el efecto total que tiene la variable sobre la medida dependiente. De manera similar, el efecto principal de la variable independiente B, el tipo de entrevistador, es el efecto que tienen los diferentes tipos de preguntas sobre la exactitud de los recuerdos.

El efecto principal de cada variable independiente es la relación total que existe entre las variables independiente y dependiente. ¿Existe una relación entre el tipo de pregunta y los errores de memoria en la variable independiente *A?* Lo sabremos si observamos las medias totales en las condiciones con preguntas sin sesgo y sesgadas. Estas medias se presentan al margen de la tabla 10.1. El porcentaje total de errores cometidos por los participantes en la condición con preguntas sesgadas es de 29.5, y el porcentaje de error de la condición con preguntas sin sesgo es de 13.0. Estas medias totales del efecto principal se obtienen cuando se promedia a todos los participantes en cada grupo, sin importar el tipo de entrevistador (experto o inexperto). Observe que la media total de 29.5, en la condición con preguntas sesgadas, corresponde al promedio de 41 en el grupo experto-confuso y de 18 en el grupo inexperto-confuso (este cálculo su-

pone un número igual de participantes en cada grupo). Se puede ver que, en general, se cometen más errores cuando las preguntas son sesgadas que cuando no están sesgadas. Las pruebas estadísticas nos permitirían determinar si se trata de un efecto principal significativo.

El efecto principal de la variable independiente *B* (tipo de entrevistador) es la relación general entre esa variable independiente, por sí misma, y la variable dependiente. En la tabla 10.1 se puede observar que la puntuación total en la condición del entrevistador experto es de 27.0, y que la puntuación general en el grupo del entrevistador inexperto es de 15.5. Por lo tanto, en general, hay más errores cuando el entrevistador es experto.

Interacciones Estos efectos principales nos indican que, en general, hay más errores cuando el entrevistador es experto y cuando las preguntas son sesgadas. También existe la posibilidad de una interacción; si la hay, deben calificarse los efectos principales de las variables independientes, debido a que una interacción entre las variables independientes indica que el efecto de una difiere en distintos niveles de la otra. Es decir, una interacción nos indica que el efecto de una variable independiente depende del nivel particular de la otra.

Podemos observar una interacción en los resultados del estudio de Smith Ellsworth. El efecto del tipo de preguntas difiere dependiendo de si el entrevistador es experto o inexperto. Cuando el entrevistador es experto, las preguntas sesgadas producen más errores (41 por ciento en la condición con preguntas sesgadas, contra el 13 por ciento en la condición con preguntas sin sesgo). Sin embargo, cuando el entrevistador es inexperto, el tipo de pregunta tiene poco efecto (18 por ciento para las preguntas sesgadas y 13 por ciento para las preguntas sin sesgo). Así, la relación entre el tipo de pregunta y los errores se comprende mejor cuando se consideran ambas variables independientes: debemos tomar en cuenta el hecho de que las preguntas sean sesgadas *y* que el entrevistador sea experto o inexperto.

Las interacciones pueden verse con facilidad cuando las medias de todas las condiciones se incluyen en una gráfica. La figura 10.3 muestra una gráfica de barras con los resultados del experimento de los testigos. Observe que se graficaron las cuatro medias. Dos barras comparan el tipo de entrevistador en la condición de las preguntas sin sesgo; también se incluye la misma comparación en la condición con las preguntas sesgadas. Podemos ver que los conocimientos del interrogador no constituyen un factor cuando se plantean preguntas sin sesgo; sin embargo, cuando la pregunta es sesgada, el entrevistador experto tiene mayores posibilidades de crear un sesgo que el inexperto.

La interacción es un concepto relativamente simple que probablemente usted utiliza todo el tiempo. Cuando decimos "depende de", generalmente indicamos que opera cierto tipo de interacción; que depende de alguna otra variable. Suponga, por ejemplo, que un amigo le pregunta si desea ir al cine. La decisión de ir puede reflejar una interacción entre dos variables: 1) ¿tiene usted un examen pronto?, y 2) ¿quién protagoniza la película? Si se aproxima un examen, usted no irá bajo ninguna circunstancia. Si no tiene ningún examen que lo

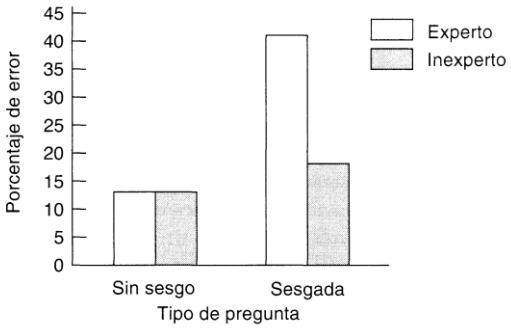


FIGURA 10.3
Interacción entre el tipo de pregunta y el tipo de entrevistador
(basado en datos de Smith y Ellsworth, 1987).

preocupe, su decisión dependerá de los actores que protagonizan la película; es decir, usted irá únicamente si participa uno de sus actores favoritos.

Trate de graficar el ejemplo del cine de la misma forma que lo hicimos con el ejemplo del testigo en la figura 10.3. La variable dependiente (ir al cine) siempre se ubica en el eje vertical, y una variable independiente en el eje horizontal. Después se dibujan barras para representar a cada uno de los niveles de la otra variable independiente. Graficar los resultados de esta forma es un método útil para visualizar las interacciones en un diseño factorial.

Diseños factoriales con manipulación y sin manipulación de variables

Un tipo común de diseño factorial incluye variables experimentales (manipuladas) y no experimentales (medidas o no manipuladas). Estos diseños, que en ocasiones se conocen como **diseños VI** X **VP** (variable independiente por variable de los participantes), permiten que los científicos investiguen cómo los distintos tipos de individuos (participantes) responden a la misma variable manipulada. Estas "variables de los participantes" son los atributos personales como el género, la edad, el grupo étnico, las características de la personalidad y la categoría clínica diagnóstica. En ocasiones, las variables de los participantes se describen como variables del sujeto o de atributo; únicamente se trata de una diferencia de términos.

El diseño VI X VP más sencillo incluye una variable independiente manipulada *con* al menos dos niveles, y una variable de los participantes con al menos dos niveles. Los dos niveles de la variable de los sujetos podrían ser dos grupos de edades diferentes, grupos con puntajes bajos y alto s en una prueba de personalidad, o grupos de hombres y mujeres. Un ejemplo de este diseño es el estudio de Furnham, Gunter y Peterson (1994). ¿Alguna vez ha intentado estudiar cuando hay una distracción, como un programa de televisión? Furnham y sus colaboradores demostraron que la habilidad para estudiar con este tipo de distracciones depende de que el individuo sea extrovertido o introvertido. La variable manipulada era la distracción. Estudiantes universitarios leyeron material en

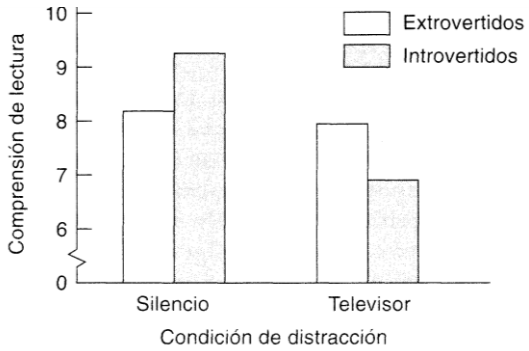
silencio, mientras escuchaban un drama televisivo. Así, se utilizó un diseño de medidas repetidas y se contrabalanceó el orden de las condiciones. Después de leer el material, los estudiantes respondieron una prueba de comprensión de la lectura. La variable de los participantes fue la extroversión: los individuos respondieron una prueba de extroversión y después fueron clasificados como extrovertidos o introvertidos. Los resultados se muestran en la figura 10.4; se observó un efecto principal de la distracción y una interacción.

En general, los estudiantes obtuvieron puntajes de comprensión más alta cuando estudiaron en silencio. Además, hubo una interacción entre la extroversión y la distracción. Sin distracción, el desempeño de los extrovertidos e introvertidos fue casi idéntica. Sin embargo, los estudiantes extrovertidos se desempeñaron mejor que los introvertidos cuando el televisor estaba encendido. Si usted es extrovertido, ¡sea más comprensivo cuando sus amigos introvertidos deseen estudiar en silencio!

Los diseños factoriales con variables independientes manipuladas y variables de los participantes, representan un método muy atractivo para estudiar numerosas preguntas de investigación interesantes. Este tipo de experimentos reconocen que una comprensión completa del comportamiento requiere que conozcamos las variables situacionales y los atributos personales de los individuos.

Interacciones y variables moderadoras

En muchos estudios, las interacciones se interpretan en términos de la operación de una **variable moderadora.** Una variable moderadora inñuye en la relación que existe entre otras dos variables (Barón y Kenny, 1986). En el estudio del jurado, podemos iniciar con una aseveración general de la relación entre el tipo de pregunta y los errores de recuerdo; por ejemplo, *las preguntas sesgadas producen mas errores que las preguntas sin sesgo.* Después podríamos aseverar cualitativamente que el tipo de entrevistador afecta esta relación: *las preguntas sesgadas dan como resultado mas errores, únicamente cuando el entrevistador parece ser experto; cuan-*



FSGURA **10.4**
Interacción en un diseño VI X VR

do se trata de un entrevistador inexperto, no hay un incremento en el número de errores. La variable del entrevistador es una variable moderadora debido a que *modera* la relación entre las otras variables. Las variables moderadoras pueden ser situaciones particulares, como en el estudio del jurado realizado por Smith y Ellsworth (1987), o pueden ser características de las personas, como en la investigación sobre la actividad de estudio de individuos extrovertidos e introvertidos.

Resultados de un diseño factorial 2 x 2

Un diseño factorial 2 X 2 incluye dos variables independientes, cada una con dos niveles. Cuando se analizan los resultados, hay varias posibilidades: 1) puede o no ocurrir un efecto principal significativo para la variable independiente *A;* 2) puede o no ocurrir un efecto principal significativo para la variable independiente *B, y* 3) puede o no ocurrir una interacción significativa entre las variables independientes.

La figura 10.5 ilustra los ocho resultados posibles de un diseño factorial 2 X 2. Para cada resultado, se incluyen las medias y gráficas de barras. Las medias representan ejemplos ideales, este tipo de resultados perfectos raras veces ocurren en investigaciones reales. No obstante, analice las gráficas para determinar por qué, en cada caso, existe o no un efecto principal para *A,* un efecto principal para *B y* una interacción *A* X *B.*

Las primeras cuatro gráficas ilustran resultados en los que no existe una interacción *A* X *B, y* las cuatro últimas describen resultados en los que sí existe dicha interacción. Cuando hay una interacción estadísticamente significativa, es necesario examinar con cuidado las medias para comprender por qué ocurrió la interacción. En algunos casos, existe una fuerte relación entre la primera variable independiente y la variable dependiente, en relación con un nivel de la segunda variable independiente; sin embargo, no existe una relación o ésta es débil, en el otro nivel de la segunda variable independiente. En otros estudios, la interacción puede indicar que una variable independiente tiene efectos opuestos en la variable dependiente, dependiendo del nivel de la segunda variable independiente. En la última gráfica de la figura 10.5 se muestra este patrón.

Las variables independientes y dependientes de las figuras no poseen etiquetas concretas para las variables. Como ejercicio, interprete cada una de las gráficas utilizando variables reales de tres experimentos hipotéticos diferentes. Esto funciona mejor si dibuja las gráficas, incluyendo las etiquetas para las variables, en una hoja de papel separada para cada experimento.

Experimento hipotético 1: efectos de la edad del acusado de un crimen *y* el tipo de sustancia utilizada durante el crimen, sobre los meses de sentencia dictados. Se encontró culpable a un hombre, de 20 o 50 años de edad, de causar un accidente de tránsito mientras se encontraba bajo la influencia del alcohol o de la mariguana.

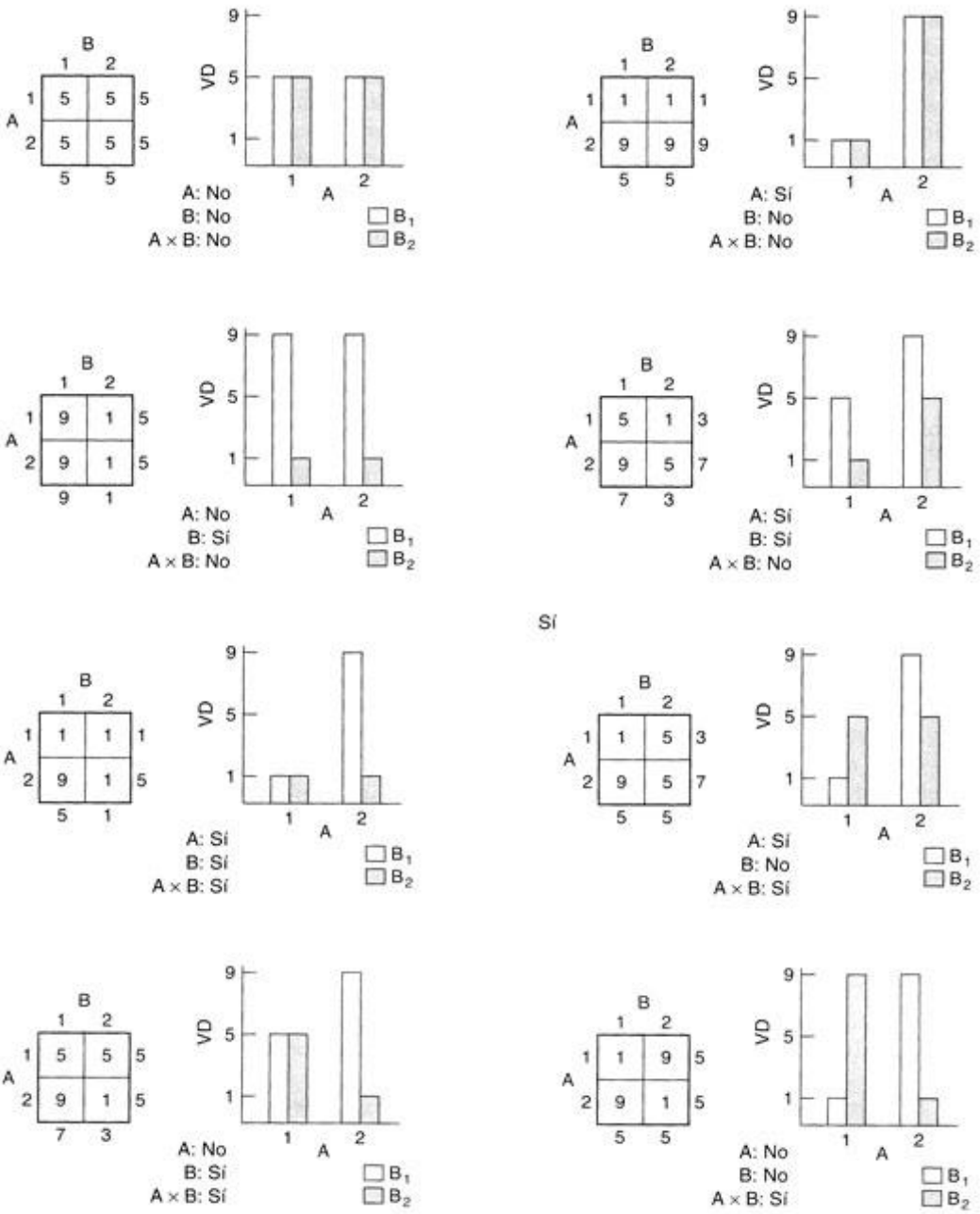


FIGURA 10.5
Resultados de un diseño factorial con dos variables independientes.

Variable independiente *A:* tipo de crimen—alcohol *vs.* mariguana Variable independiente *B:* edad del acusado—20 *vs.* 50 años de edad Variable dependiente: meses de sentencia (de cero a diez meses)

Experimento hipotético 2: efectos del género y la violencia sobre el recuerdo de anuncios. Los participantes (hombres y mujeres) observaron una filmación violenta o no violenta en una pantalla de computadora. Después, durante los siguientes tres minutos, se les pidió que leyeran anuncios impresos de ocho productos diferentes. La variable dependiente fue el número de anuncios recordados de manera correcta.

Variable independiente *A:* exposición a la violencia: filmación no violenta *vs.* filmación violenta

Variable independiente *B:* género del participante: hombre *vs.* mujer

Variable dependiente: número de anuncios recordados (de cero a ocho)

Experimento hipotético 3: diseñe su propio experimento con dos variables independientes y una dependiente.

Interacciones y efectos principales simples

En un diseño factorial, el procedimiento denominado análisis de varianza se utiliza para evaluar la significación estadística de los efectos principales y de la interacción. Cuando existe una interacción significativa, es necesario evaluar de manera estadística las medias individuales. Si observa la figura 10.3 nuevamente, encontrará una clara interacción. Cuando la interacción es significativa, el siguiente paso es analizar los **efectos principales simples.** Un análisis de los efectos principales simples examina las diferencias entre las medias, en *cada nivel* el la variable independiente. Recuerde que el efecto principal de una variable independiente es un promedio de los niveles de la otra variable independiente; con los efectos principales simples, los resultados se analizan como si tuviéramos experimentos separados en cada nivel de la otra variable independiente.

Efecto principal simple del tipo de entrevistador En la figura 10.3 podemos observar el efecto principal simple del tipo de entrevistador. Éste nos indica si la diferencia entre el entrevistador experto y el no experto es significativa cuando la pregunta 1) no tiene sesgo y 2) es sesgada. En este caso, el efecto principal simple del tipo de pregunta es significativo cuando ésta es sesgada (con medias de 41 *vs.* 18), aunque el efecto principal simple del tipo de pregunta no es significativo cuando ésta no tiene sesgo (con medias de 13 y 13).

Efecto principal simple del tipo de pregunta También podemos examinar el efecto principal simple del tipo de pregunta; aquí, compararíamos las dos preguntas

cuando el entrevistador es 1) experto y 2) inexperto. El efecto principal simple que le interesará más dependerá de las predicciones que haya planteado cuando diseñó el estudio. Los procedimientos estadísticos exactos no nos interesan ahora; lo más importante es que, cuando exista una interacción significativa en un diseño factorial, debe examinar el patrón de resultados con todas las medias.

Procedimientos de asignación y diseños factoriales

Los aspectos que se estudiaron en el capítulo 8, sobre la asignación de los participantes a las distintas condiciones, pueden generalizarse a los diseños factoriales. Hay dos formas básicas para asignar a los participantes a las distintas condiciones: 1) en un diseño de grupos independientes, se asignan distintos participantes a cada una de las condiciones del estudio; 2) en un diseño de medidas repetidas, los *mismos* individuos participan en todas las condiciones del estudio. Estos dos tipos de procedimientos de asignación tienen implicaciones en el número de participantes que se requieren para completar el experimento. Podemos ilustrar este hecho al observar un diseño factorial 2 x 2. El diseño puede ser exclusivamente de grupos independientes, exclusivamente de medidas repetidas o un **diseño factorial mixto,** es decir, una combinación de ambos.

Grupos independientes En un diseño factorial 2 X 2, existen cuatro condiciones. Si deseamos un diseño que sea exclusivamente de grupos independientes, se asigna un grupo distinto de participantes a cada una de las cuatro condiciones. El estudio de Smith y Ellsworth (1987) sobre el reporte de testigos y el de Petty, Cacioppo y Goldman (1981) acerca de los efectos de la credibilidad del comunicador y el involucramiento personal sobre el cambio de actitudes, descritos en el capítulo 9, son ejemplos de diseños factoriales con distintos individuos en cada condición. Suponga que usted planea un diseño 2 x 2 y que desea incluir a diez participantes en cada condición; necesitará un total de 40 participantes *diferentes, tal* como se muestra en la primera tabla de la figura 10.6.

Medidas repetidas En un procedimiento que utiliza únicamente medidas repetidas, los mismos individuos participan en *todas* las condiciones. Suponga que usted planea un estudio sobre los efectos de la mariguana, similar al realizado por Wilson, Ellinwood, Mathew y Johnson (1994), descrito en el capítulo 8: un factor es la mariguana (tratamiento de mariguana *vs.* placebo), y el otro factor es la dificultad de una tarea (fácil *vs.* difícil). En un diseño 2 X 2 que utiliza únicamente medidas repetidas, cada individuo participaría en todas las condiciones, realizando las tareas fáciles y las difíciles bajo los dos tratamientos. Si usted quisiera que hubiese diez participantes en cada condición, necesitaría un total de diez sujetos, como se observa en la segunda tabla de la figura 10.6. Este diseño nos brinda un ahorro considerable en la cantidad de participantes requeridos. Sin embargo, para decidir si utiliza un procedimiento de asignación de medidas repetidas, el investigador tendría que considerar las desventajas de estos diseños.

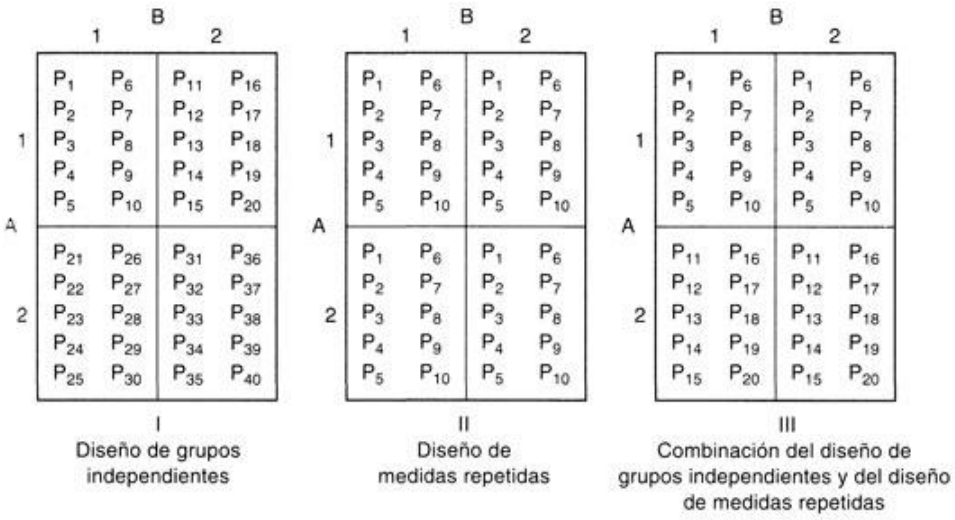


FIGURA 10.6
Número de participantes requeridos para tener diez observaciones en cada condición.

Diseño factorial mixto con el uso de asignación combinada El estudio de Furnham, Gunter y Peterson (1994), sobre la distracción de la televisión y la extraversión, ejemplifica el uso de los procedimientos de grupos independientes y de medidas repetidas en un diseño factorial mixto. La variable de los participantes, la extraversión, es una variable de grupos independientes. La distracción es una variable de medidas repetidas; todos los participantes estudiaron con distracción y en silencio. La tercera tabla de la figura 10.6 muestra la cantidad de participantes que se requieren para tener diez en cada condición, en un diseño factorial mixto 2 X 2. En esta tabla, la variable independiente *A* constituye una variable de grupos independientes. Se asignan diez participantes al nivel 1 de esta variable independiente, y otros diez se asignan al nivel 2. Sin embargo, la variable independiente *B* es una variable de medidas repetidas. Los diez participantes asignados a A_1 reciben ambos niveles de la variable independiente *B.* De manera similar, los otros diez participantes, asignados a A_2, reciben ambos niveles de la variable *B.* Aquí, se requiere un total de 20 participantes.

Incremento del número de niveles de una variable independiente

El diseño 2 x 2 es el diseño factorial más simple. Con este diseño básico, el investigador puede elaborar experimentos cada vez más complejos. Una forma de incrementar la complejidad es aumentar el número de niveles de una o más variables independientes. Por ejemplo, un diseño 2 X 3 incluye dos variables independientes: la variable independiente *A* tiene dos niveles, y la variable inde-

pendiente *B* incluye tres niveles. Por lo tanto, el diseño 2 X 3 tiene seis condiciones. La tabla 10.2 muestra un diseño factorial 2 X 3 con las variables independientes de la dificultad de la tarea (fácil, difícil) y el nivel de ansiedad (bajo, moderado, alto); la variable dependiente es el desempeño en la tarea. Las cifras que aparecen en cada una de las seis celdas del diseño indican el desempeño promedio del grupo. Las medias totales al margen muestran los efectos principales de cada una de las variables independientes. Los resultados en la tabla 10.2 indican un efecto principal en la dificultad de la tarea, ya que el puntaje *total* del desempeño en el grupo de la tarea fácil es más alto que la media de la tarea difícil. Sin embargo, no existe un efecto principal de la ansiedad, debido a que el promedio del desempeño es igual en cada uno de los tres grupos de ansiedad. ¿Existe una interacción entre la dificultad de la tarea y la ansiedad? Observe que un incremento en la cantidad de la ansiedad tiene el efecto de mejorar el desempeño en la tarea fácil, pero *empeora* el desempeño en la tarea difícil. El efecto de la ansiedad difiere, dependiendo de la dificultad de la tarea, por lo cual existe una interacción.

La interacción puede verse fácilmente en una gráfica. Las figuras anteriores de este capítulo han incluido gráficas de barras; la figura 10.7 es una gráfica lineal. Los datos de una gráfica de barras y de una gráfica lineal pueden ser los mismos, pero la representación visual difiere de cierta forma. Se dibujan líneas en lugar de barras para mostrar los hallazgos en cada variable independiente. En la figura 10.7, una línea muestra el efecto de la ansiedad en la tarea fácil, y la segunda representa el efecto de la ansiedad en la tarea difícil. Las gráficas lineales se utilizan cuando la variable independiente, representada en el eje horizontal, es de tipo cuantitativo. Los niveles de la variable independiente son cantidades crecientes de la variable.

Incremento del número de variables independientes en un diseño factorial

También podemos incrementar el número de variables en el diseño. Un diseño factorial 2 X 2 X 2 incluye tres variables, cada una con dos niveles. Por lo tanto,

TABLA 10.2 Diseño factorial 2 × 3.

	Nivel de ansiedad			
Dificultad de la tarea	Bajo	Moderado	Alto	Medias totales (efecto principal)
Fácil	4	7	10	7.0
Difícil	7	4	1	4.0
Medias totales (efecto principal)	5.5	5.5	5.5	

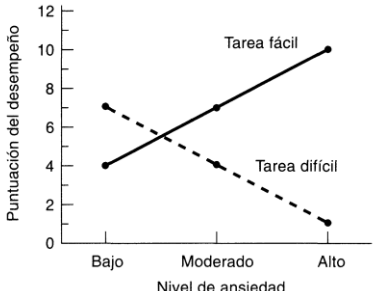


FIGURA 10.7
Gráfica lineal de los datos de un diseño factorial 3 (nivel de ansiedad) × 2 (dificultad de la tarea).

este diseño tiene 8 condiciones. En un diseño 2 X 2 X 3 , existen 12 condiciones; en un diseño 2 X 2 X 2 X 2 , existen 16. La regla para elaborar diseños factoriales siempre es la misma.

En la tabla 10.3 se muestra un diseño factorial 2 X 2 X 2 . Las variables independientes son: 1) método de enseñanza (conferencia, debate), 2) tamaño de la clase (10, 40) *y* 3) género del estudiante (hombre, mujer). Observe que el género no es una variable manipulada y que las otras dos sí lo son. La variable dependiente es el desempeño en una prueba estándar.

Observe también que un diseño 2 X 2 X 2 puede verse como dos diseños 2X2, uno para los hombres y otro para las mujeres. El diseño produce efectos principales para cada una de las tres variables independientes. Por ejemplo, las medias totales del método de conferencia se obtienen considerando a todos los participantes que experimentaron este método, sin importar el tamaño de la clase o el género. De manera similar, la media del método de debate deriva de todos los

TABLA 10.3 Diseño factorial 2 X 2 X 2 .

	Método de enseñanza	
Tamaño de la clase	10	40
	Hombre	
Conferencia		
Debate		
	Mujer	
Conferencia		
Debate		

participantes en esta condición. Después, las *dos* medias se comparan para ver si existe un efecto principal significativo: *¿en general,* es mejor un método que otro?

El diseño también nos permite observar interacciones. En el diseño 2 X 2 X 2 podemos encontrar la interacción entre: 1) el método y el tamaño de la clase, 2 el método y el género, y 3) el tamaño de la clase y el género. También podemos ver una interacción de tres factores que incluye a las tres variables independientes. Aquí debemos determinar si la naturaleza de la interacción entre dos de las variables difiere dependiendo del nivel particular de la otra variable. Las interacciones de tres factores son bastante complicadas; por fortuna, no se encontrarán muchas de éstas cuando explore investigaciones del comportamiento.

En ocasiones, los estudiantes se ven tentados a incluir en un estudio todas las variables independientes que se les ocurren. El problema de esto es que el diseño puede volverse innecesariamente complejo y requerir cantidades enormes de participantes. El diseño expuesto previamente incluía ocho grupos; un diseño 2 X 2 X 2 X 2 incluye 16 grupos; al añadir otra variable independiente con dos niveles, se requerirían 32 grupos. Asimismo, cuando hay más de tres o cuatro variables independientes, muchas de las condiciones particulares que se producen por la combinación de tantas variables no tendrían sentido o no podrían ocurrir en circunstancias naturales.

Los diseños hasta ahora descritos utilizan la misma lógica para determinar si la variable independiente causó, en realidad, un cambio en la medida de la variable dependiente. En el próximo capítulo estudiaremos diseños alternos que utilizan procedimientos un poco distintos para examinar la relación entre variables independientes y dependientes.

Términos de estudio

Diseño factorial
Diseño factorial mixto
Diseño VI x VP
Efecto principal
Efecto principal simple
Interacción
Variable moderadora

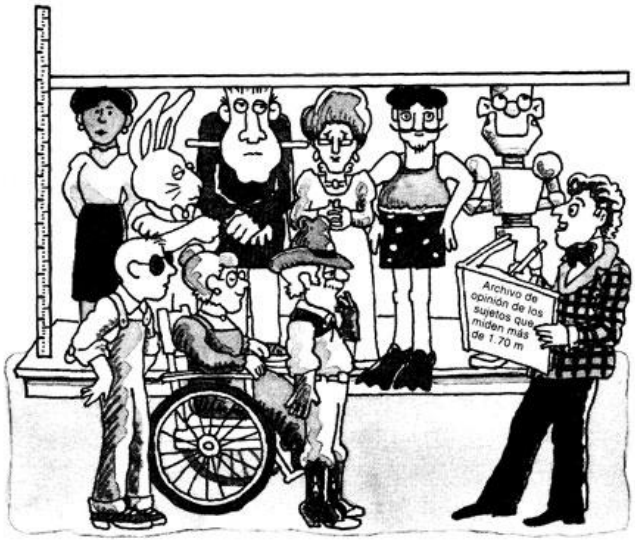
Preguntas de repaso

1. ¿Por qué un investigador incluiría más de dos niveles de la variable independiente en un experimento?
2. ¿Qué es un diseño factorial? ¿Por qué los investigadores utilizan diseños factoriales?
3. ¿Qué son los efectos principales en un diseño factorial? ¿Qué es una interacción?
4. Describa un diseño factorial VI x VP.
5. Identifique el número de condiciones en un diseño factorial, con base en el conocimiento del número de variables independientes y el número de niveles de cada variable independiente.

Actividades

1. En un estudio realizado por Chaiken y Pliner (1987), los participantes de la investigación leyeron un "diario de alimentación" de un hombre y uno de una mujer. La información en el diario indicaba que la persona consumía grandes o pequeñas cantidades de alimento. Después de leer esta información, los participantes calificaron la feminidad y masculinidad del individuo.
   a) Identifique el diseño de este experimento.
   b) ¿Cuántas condiciones hay en el experimento?
   c) Identifique la(s) variable(s) independiente(s) y la(s) variable(s) dependiente(s).
   d) ¿Existe una variable de los participantes en este experimento? Si es así, identifíquela. Si no es así, ¿podría sugerir una variable de los participantes que podría incluirse?
2. Chaiken y Pliner reportaron los siguientes puntajes medios de feminidad (cantidades más altas indican una mayor feminidad): hombre-pocos alimentos (2.02), hombre-muchos alimentos (2.05), mujer-pocos alimentos (3.90), y mujer-muchos alimentos (2.82). Suponga que existe el mismo número de participantes en cada condición.
   a) ¿Existe algún efecto principal?
   b) ¿Existe una interacción?
   c) Haga una gráfica de las medias.
   d) Describa los resultados en un párrafo breve.
3. Utilice revistas científicas de psicología recientes y encuentre un ejemplo de un diseño de grupos independientes 2 X 2 . Identifique cada factor y sus niveles. Encuentre otro experimento que ejemplifique un diseño de dos factores, con medidas repetidas para una o más variables; identifique si se trata de un procedimiento exclusivamente de medidas repetidas o de un diseño factorial mixto. Localice cada factor y sus niveles. ¿Existe una interacción? Si es así, descríbala.
4. Suponga que desea tener 15 participantes en cada condición de un experimento que utiliza un diseño factorial 3X3. ¿Cuántos participantes *distintos* necesita para: *a)* una asignación exclusivamente de grupos independientes, *b)* una asignación exclusivamente de medidas repetidas
   y *c)* un diseño factorial mixto con asignación de grupos independientes y variables de medidas repetidas?

11



Diseños cuasiexperimentales y experimentales de caso único

Evaluación de programas

Diseños cuasiexperimentales

Diseño de un grupo con sólo postest Diseño de un grupo con pretest-postest

Historia

Maduración

Prueba

Problemas de instrumentación

Regresión hacia la media Diseño de grupo control no equivalente Diseño de grupo control no equivalente con pretest-postest

Diseño de series de tiempo interrumpidas Diseño de series de control

Diseños experimentales de caso único

Diseños reversibles

Diseños de línea base múltiple

Replicación de los diseños de caso único

Términos de estudio

Preguntas de repaso

Actividades

En el diseño experimental clásico, descrito en el capítulo 8, los participantes se asignan de manera aleatoria a las condiciones de la variable independiente, y luego se mide la variable dependiente. Después se comparan las respuestas de la medida dependiente para determinar si la variable independiente tuvo algún efecto. Puesto que todas las demás variables se mantienen constantes, las diferencias en la variable dependiente se deben al efecto de la variable independiente. Este diseño tiene una alta validez interna, es decir, tenemos mucha confianza en que la variable independiente causó las respuestas observadas en la variable dependiente. Con frecuencia usted se encontrará este diseño experimental cuando lea investigaciones en las ciencias del comportamiento. Sin embargo, se han creado otros diseños de investigación para enfrentar problemas especiales. Este capítulo se enfoca en los diseños de investigación cuasiexperimental y de caso único; métodos de investigación que representan interesantes retos para determinar la validez interna.

Como señalamos en el capítulo 1, los científicos con frecuencia investigan preguntas aplicadas y realizan investigación de evaluación. Los diseños cuasiexperimentales surgieron de la necesidad de llevar a cabo investigaciones aplicadas en ambientes en los que no pueden lograrse las características de control de los diseños experimentales completos. En un **diseño cuasiexperimental** existe la necesidad de evaluar el efecto de una variable independiente sobre una dependiente. No obstante, es más difícil hacer inferencias causales debido a que los cuasiexperimentos carecen de las características de control de los experimentos completos. En este capítulo examinamos varios diseños cuasiexperimentales que pueden utilizarse en situaciones donde no es posible realizar experimentos completos. Aun cuando los diseños cuasiexperimentales pueden emplearse para estudiar numerosos problemas de investigación, es útil considerarlos en el contexto de la investigación de evaluación de programas.

EVALUACIÓN DE PROGRAMAS

En el capítulo 1 se describió brevemente la **evaluación de programas.** Se trata de la investigación de programas que se proponen y llevan a cabo para lograr algún efecto positivo en un grupo de individuos. Este tipo de programas puede aplicarse en escuelas, ambientes de trabajo o comunidades enteras. Un ejemplo de las escuelas es el programa DARE (Drug Abuse Resistance Education), diseñado para reducir el consumo de drogas. Este proyecto se aplica en conjunto con departamentos locales de policía y se ha vuelto muy popular desde su creación a principios de los años ochenta. La evaluación de programas aplica muchas de las estrategias de investigación para estudiar este tipo de programas.

Donald Campbell (1969) motivó una cultura de la evaluación en la que todos los programas deben ser analizados de forma honesta para determinar si son eficaces. Así, el enfoque inicial de la investigación de evaluación fue la "evaluación de los resultados": ¿el programa dio los resultados positivos para los que fue creado (por ejemplo, la reducción del consumo de drogas, calificaciones

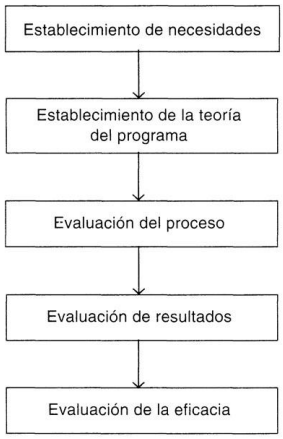


FIGURA 11.1
Fases de la investigación de evaluación de programas.

más altas, menor ausentismo o menor reincidencia)? Sin embargo, conforme el campo de la evaluación de programas ha avanzado, desde el artículo de Campbell publicado en 1969, la investigación evaluativa abarca cada vez más que la simple evaluación de resultados (Rossi, Freeman y Lipsey, 1999).

Rossi y sus colaboradores (1999) identificaron cinco tipos de evaluaciones; cada una intenta responder una pregunta distinta acerca del programa. En la figura 11.1 se muestran las cinco fases del proceso de evaluación. La primera es la evaluación de necesidades. Los estudios de *evaluación de necesidades* preguntan si, de hecho, existen problemas que necesiten estudiarse en una población específica. Por ejemplo, ¿existe un abuso de drogas por parte de los niños y adolescentes en la comunidad? Si es así, ¿qué tipo de drogas se están consumiendo? ¿Qué servicios necesitan más los individuos sin hogar? ¿Los criminales juveniles reincidentes tienen problemas personales y familiares particulares que podrían abordarse por medio de un programa de intervención? Una vez que se ha detectado una necesidad, se pueden planear programas para cubrirla. Los datos sobre el establecimiento de necesidades pueden provenir de encuestas, entrevistas y datos estadísticos obtenidos por diversos organismos públicos.

El segundo tipo de pregunta de evaluación de un programa se refiere a su teoría. Después de identificar las necesidades, se puede diseñar un programa para resolverlas. Rossi y sus colaboradores (1999) hacen énfasis en que el programa debe basarse en supuestos válidos acerca de las causas de los problemas y en sus fundamentos. La *evaluación de la teoría del programa* puede incluir la colaboración de investigadores, proveedores de servicios y clientes potenciales del programa para determinar que éste, de hecho, cubra las necesidades de la población específica de forma apropiada. Rossi y sus colaboradores describen un estudio que

evaluó las necesidades de hombres y mujeres sin hogar en la ciudad de Nueva York (Herman, Struening y Barrow, 1994). Las necesidades generales más importantes implicaban ayuda para encontrar un lugar para vivir, un empleo y mejorar sus habilidades para trabajar. Los hombres, en particular, necesitaban ayuda para superar sus problemas con el alcohol y las drogas, para aprender a manejar el dinero y para tener una buena relación con los demás. Las mujeres requerían ayuda para problemas de salud y de tipo médico. Un programa que esté diseñado para cubrir estas necesidades debe tomar en cuenta esta clase de información y tener un fundamento con respecto de la forma en que los individuos sin hogar obtendrán en realidad beneficios del programa.

El tercer tipo de pregunta en la evaluación de programas es la *evaluación del proceso* o monitoreo del programa. Una vez que el programa se ha iniciado, el investigador que hace la evaluación realiza verificaciones para determinar si está alcanzando a la población específica, si está atrayendo a suficientes clientes y si el personal está proporcionando los servicios planeados. En ocasiones, el personal no ha recibido un entrenamiento pertinente o los servicios se están ofreciendo en una localidad inadecuada o difícil de encontrar. En resumen, el investigador desea asegurarse de que en el programa se esté haciendo lo que se supone debe hacerse. Esta investigación es sumamente importante, ya que no deseamos concluir que un programa es ineficiente si, de hecho, es la *implementación* del programa lo que no está funcionando. La investigación puede hacerse con cuestionarios y entrevistas, estudios observacionales y análisis de los registros obtenidos por el personal del programa.

La cuarta pregunta se refiere a la *evaluación de resultados* o evaluación del impacto. ¿Se están cumpliendo los objetivos planeados en el programa? ¿Se está logrando la meta: reducir el consumo de drogas, aumentar el alfabetismo, disminuir la reincidencia de crímenes en los jóvenes o desarrollar las habilidades para el trabajo? Para determinar esto, el evaluador debe diseñar una forma para medir los resultados y después estudiar la trascendencia del programa sobre el resultado medido. Es necesario saber cómo son los participantes del programa y cómo serían si no hubiesen participado en éste. De manera ideal, se realizaría un diseño experimental con asignación aleatoria a las condiciones para responder preguntas acerca de los resultados. Sin embargo, otros métodos de investigación, como los diseños cuasiexperimental y de caso único, descritos en este capítulo, son muy útiles para evaluar la repercusión de un programa de intervención.

La pregunta final de la evaluación de un programa se refiere a la *evaluación de la eficacia.* Una vez que se ha demostrado que un programa ejerce los efectos buscados, los investigadores deben determinar si "vale la pena". Se deben ponderar los costos del programa frente a sus beneficios. Además, los investigadores deben determinar si podría darse un mejor uso a los recursos utilizados para realizar el programa.

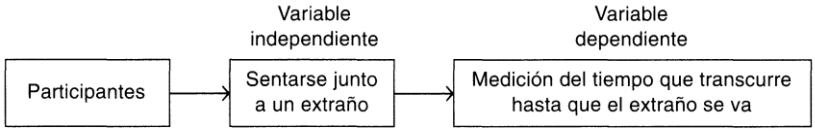
Como se señaló, la investigación de evaluación puede llevarse a cabo por medio de toda s las aproximaciones metodológicas estudiadas en los capítulos anteriores: encuestas, observación naturalista y experimentos completos. No obstante, este tipo de investigación generalmente debe realizarse en ambientes y

bajo circunstancias en las que no se puede aplicar un diseño experimental completo. En situaciones como éstas, debe utilizarse una alternativa al diseño experimental completo, como el diseño cuasiexperimental. A continuación revisaremos únicamente algunos de los diseños cuasiexperimentales más utilizados. Para una explicación más detallada, véase Campbell (1968, 1969); Shadish, Cook y Campbell (2002); y Campbell y Stanley (1966). Después examinaremos los diseños de caso único que pueden emplearse para la evaluación de programas y en muchas otras áreas de investigación.

DISEÑOS CUASIEXPERIMENTALES

Diseño de un grupo con sólo postest

Suponga que desea investigar si el hecho de sentarse cerca de un extraño causa que éste se aleje. Usted podría sentarse junto a una cantidad de extraños y medir el número de segundos que transcurren antes de que se alejen. Su diseño se vería como sigue:



Ahora, suponga que el promedio de tiempo que transcurre antes de que las personas se alejen es de 9.6 segundos. Desafortunadamente, este hallazgo no puede interpretarse. Usted no sabe si hubiesen permanecido más tiempo en caso de que usted no se hubiera sentado, o si hubiesen permanecido durante 9.6 segundos, de cualquier manera. Incluso, es posible que se fueran antes si usted no se hubiera sentado, ¡tal vez usted les agradó!

Este **diseño de un grupo con sólo postest,** denominado "estudio de caso de una sola aplicación" por Campbell y Stanley (1966), carece del elemento crucial de un experimento completo: un grupo control o de comparación. Debe haber algún tipo de condición de comparación que le permita interpretar sus resultados. El diseño de grupo control con sólo postest sin un grupo de comparación tiene graves deficiencias en el contexto de diseñar un experimento con validez interna, que permita hacer inferencias causales acerca de los efectos de una variable independiente sobre una variable dependiente.

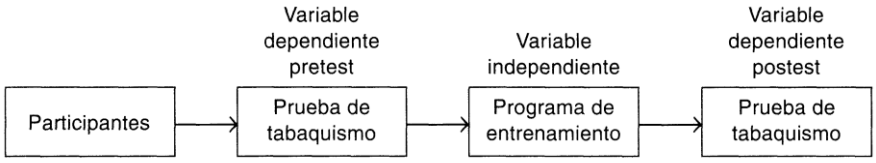
Quizá se pregunte si este diseño realmente se llega a utilizar. De hecho, se puede considerar que este tipo de diseño se utiliza como evidencia de la eficacia de un programa. Por ejemplo, los empleados de una empresa podrían participar en una sesión informativa de cuatro horas sobre procedimientos de emergencia. Al final del programa, ellos responden una prueba de conocimientos en la que

su puntaje promedio es de 90 por ciento. Luego, este resultado se utiliza para concluir que el programa es exitoso para entrenar empleados. El problema, por supuesto, es que el puntaje pudo haber sido igual, más bajo e incluso más alto sin la aplicación del programa. La razón por la que en ocasiones se aceptan resultados como éstos, es que podemos tener una idea implícita de la forma en que se desempeñaría un grupo control. Por desgracia, necesitamos los datos de comparación.

Diseño de un grupo con pretest-postest

Una forma de obtener una comparación es medir a los participantes antes de la manipulación (un pretest) y nuevamente al final (un postest). Entonces, se puede calcular un índice de cambio del pretest al postest. Aun cuando este **diseño de un grupo con pretest-postest** parece bueno, tiene algunos problemas importantes.

Como ejemplo, suponga que usted desea probar la hipótesis de que un programa de entrenamiento en relajación causará una reducción del tabaquismo. Si utilizara el diseño de un grupo con pretest-postest, seleccionaría a un grupo de personas que fuman, aplicaría una prueba de tabaquismo, les daría el entrenamiento de relajación y después reaplicaría la prueba de tabaquismo. Su diseño se vería como sigue:



Si usted realmente encuentra una reducción del tabaquismo, no podría asumir que los resultados se deben al programa de entrenamiento en relajación. Este diseño no tomó en cuenta varias explicaciones alternas: la historia, la maduración, las pruebas, problemas de instrumentación y la regresión hacia la media.

Historia La *historia* se refiere a cualquier suceso que ocurre entre la primera y la segunda prueba, pero que no forma parte de la manipulación. Cualquier acontecimiento de éstos se confunde con la manipulación. Por ejemplo, suponga que una persona famosa muere de cáncer pulmonar entre la primera y la segunda prueba. Este hecho, y no el entrenamiento de relajación, podría ser el responsable de la reducción del tabaquismo. Obviamente, el ejemplo de la muerte del individuo famoso es extremo y quizás improbable. Sin embargo, los **efectos de historia** pueden provenir de cualquier suceso confuso que ocurra al mismo tiempo que la manipulación experimental.

Maduración Las personas cambian con el tiempo. En un corto periodo pueden aburrirse, fatigarse, tener hambre e incluso mejorar su desempeño; después de

un largo periodo, los niños se vuelven más coordinados y analíticos. A cualquier cambio que ocurra de forma sistemática con el paso del tiempo se le llama **efecto de maduración.** La maduración puede representar un problema en el ejemplo de la reducción del tabaquismo, si las personas se preocupan más por su salud conforme crecen. Cualquier factor de tiempo como éste podría causar un cambio del pretest al postest. Si esto sucede, usted podría atribuir erróneamente el cambio al tratamiento y no a la maduración.

Prueba La prueba se convierte en un problema si el simple hecho de responder el pretest cambia el comportamiento del participante. Por ejemplo, medir el tabaquismo podría requerir que las personas llevaran un diario en el que anotaran cada cigarrillo fumado durante el día. El simple hecho de llevar un control del tabaquismo podría ser suficiente para causar una reducción en la cantidad de cigarrillos que una persona fuma. Así, la reducción detectada en el postest podría ser el resultado del pretest y no del programa en sí. En otros contextos, la aplicación del pretest podría sensibilizar a las personas de los propósitos del experimento o mejorar la habilidad que se está estudiando. Nuevamente, el experimento no tendría validez interna.

Problemas de instrumentación En ocasiones, las características básicas del instrumento de medición cambian con el tiempo; a esto se le llama **problema de instrumentación.** Considere fuentes de problemas de instrumentación cuando son observadores humanos los que miden las conductas. Con el tiempo, un observador puede volverse más hábil, fatigarse o cambiar los estándares en los que basa sus observaciones. En el ejemplo del tabaquismo, los participantes podrían estar muy motivados para registrar todos los cigarrillos fumados durante el pretest, cuando la tarea es novedosa e interesante; pero al momento del postest pueden estar cansados de la tarea y olvidar registrar un cigarrillo. Este tipo de problema de instrumentación reflejaría una aparente reducción en el consumo de cigarrillos.

Regresión hacia la media Llamada en ocasiones *regresión estadística,* la **regresión hacia la media** puede presentarse cuando se selecciona a los participantes con base en la obtención de una puntuación extremadamente baja o alta en una variable. Cuando se prueban nuevamente, sus puntajes tienden a cambiar en dirección de la media. Los puntajes extremadamente altos pueden disminuir (cerca de la media) y los puntajes extremadamente bajos tienden a elevarse (nuevamente, cerca de la media).

La regresión hacia la media constituiría un problema en el experimento del tabaquismo si los participantes se seleccionaran porque inicialmente resultaron ser fumadores empedernidos. Al elegir para el programa a las personas con los puntajes más altos en el pretest, el investigador podría seleccionar a muchos participantes que, por cualquier razón, fumaron mucho más de lo común en el momento específico en que se administró la prueba. Los individuos que fumaron más de lo común probablemente fumen menos cuando se vuelva a medir su tabaquismo. Si, entonces, comparamos la cantidad total de cigarrillos antes y

después del programa, parecerá que fuman menos. La explicación alterna es que la reducción del tabaquismo se debe a una regresión estadística, más que a un efecto del programa.

La regresión hacia la media ocurre siempre que se reúne un conjunto de puntajes extremos, obtenidos en un momento del tiempo, y se comparan con puntajes obtenidos en otro momento. En realidad, el problema radica en la confia-bilidad de la medida. En el capítulo 5 vimos que cualquier medida refleja un puntaje verdadero más el error de medición. Si existe una confiabilidad perfecta, las dos medidas serán iguales (si no sucede algo que disminuya o incremente los puntajes). Si la medida del tabaquismo es perfectamente confiable, una persona que reporta fumar 20 cigarrillos hoy, informará que fuma 20 cigarrillos dentro de dos semanas. Sin embargo, si las dos medidas no son perfectamente confiables y hay un error de medición, la mayoría de los puntajes se acercará al puntaje verdadero, aunque algunos serán altos y otros bajos. Así, un fumador con un puntaje verdadero de 20 cigarrillos por día, podría fumar *cinco* cigarrillos algunas veces y otras 35; no obstante, la mayoría de las veces el número se acercará a 20. Otro fumador podría tener una puntuación verdadera de 35, pero en ocasiones fumará 20 y en otras hasta 50; nuevamente, la mayoría de las veces el número se acercará más al puntaje verdadero que a los extremos. Ahora suponga que selecciona a dos personas que dicen haber fumado 35 cigarrillos el día anterior, y que incluye a ambas en el grupo; usted eligió a la primera persona en un día poco común y a la segunda en un día muy común. Cuando mida a estas dos personas dos semanas después, probablemente la primera reportará que fuma alrededor de 20 cigarrillos y la segunda alrededor de 35. Si promedia a las dos, aparentemente habrá una reducción general del tabaquismo.

¿Qué sucede si la medida es perfectamente confiable? En este caso, la persona con un puntaje verdadero de 20 cigarrillos siempre reportaría esta cantidad y, por lo tanto, no sería incluida en el grupo de los grandes fumadores (35+). Únicamente las personas con puntajes verdaderos de 35 o más participarían en el grupo, y cualquier reducción del tabaquismo estaría causada por el programa de tratamiento. Lo importante aquí es que la regresión hacia la media constituye un problema si existe un error de medición.

La regresión estadística también ocurre cuando tratamos de explicar sucesos de "la vida real". Los columnistas de deportes con frecuencia se refieren a la maldición que aqueja al atleta que aparece en la portada de *Sports Illustrated*. El desempeño de varios atletas ha decaído considerablemente después de ser el tema de las historias cubiertas por esta revista. Aun cuando estas historias podrían causar un menor desempeño (quizás la notoriedad causa nerviosismo y menor concentración), la regresión estadística también es una explicación probable. Se selecciona a un atleta para la portada de la revista cuando se está desempeñando en un nivel excepcionalmente alto; el principio de la regresión hacia la media establece que un desempeño muy alto tiende a deteriorarse. Podríamos asegurarnos de esto si *Sports Illustrated* también incluyera historias de atletas que están pasando una mala racha, ¡y esto se convirtiera en un buen augurio para ellos!

Todos estos problemas pueden eliminarse por medio del uso de un grupo control apropiado. Un grupo que no recibe el tratamiento experimental constituye un control adecuado para los efectos de la historia, la regresión estadística, etc. Por ejemplo, los acontecimientos históricos externos tendrían el mismo efecto en el grupo experimental que en el grupo control. Si el grupo experimental difiere del grupo control en la medida dependiente administrada después de la manipulación, la diferencia entre los dos grupos puede atribuirse a los efectos de la manipulación experimental.

Para crear un grupo control, los participantes en la condición experimental y la condición control deben ser equivalentes. Si los participantes en los dos grupos difieren *antes* de la manipulación, probablemente también diferirán *después* de la manipulación. El siguiente diseño ilustra este problema.

Diseño de grupo control no equivalente

El **diseño de grupo control no equivalente** emplea un grupo control separado, pero los participantes en las dos condiciones —el grupo experimental y grupo control— no son equivalentes. Las diferencias se convierten en una variable de confusión que proporciona una explicación alterna para los resultados. Este problema, llamado **diferencias de selección,** generalmente ocurre cuando los participantes que conforman los dos grupos en el experimento son elegidos de grupos naturales ya existentes. Si el programa de entrenamiento en relajación se estudia con un diseño de grupo control no equivalente, el diseño será como sigue:

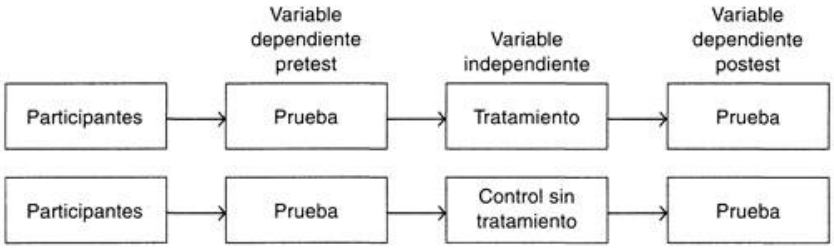
A los participantes del primer grupo se les aplica la medida de frecuencia de fumar después de completar el entrenamiento de relajación. Las personas del segundo grupo no participan en programa alguno. En este diseño, el investigador no controla en qué grupo participan los individuos. Suponga, por ejemplo, que el estudio se realiza en un departamento de una gran empresa. Se identifica a todos los empleados que fuman y se reclutan para participar en el programa de entrenamiento. Aquellos que participan de manera voluntaria en el programa se incluyen en el grupo experimental, y el grupo control se compone de los fumadores que no se inscribieron en el entrenamiento. El problema de las diferencias de selección surge debido a que los fumadores que eligen participar pueden diferir de alguna manera importante de aquellos que no lo hicieron.

Por ejemplo, tal vez de antemano sean fumadores ligeros, comparados con los demás, y tengan mayor confianza en que un programa puede ayudarlos. Si es así, cualquier diferencia entre los grupos, con respecto de la medida del tabaquismo, reflejaría diferencias previas más que los efectos del entrenamiento en relajación.

Es importante señalar que el problema de las diferencias de selección surge en este diseño, incluso cuando el investigador aparentemente ha manipulado con éxito la variable independiente, utilizando dos grupos similares. Por ejemplo, un investigador podría decidir que todos los fumadores del área de ingeniería de una empresa participaran en el programa de entrenamiento de relajación, y que los fumadores que trabajan en el área de mercadotecnia constituyeran el grupo control. El problema aquí, por supuesto, es que los fumadores de las dos áreas podrían tener patrones de tabaquismo distintos *desde antes* del programa de relajación.

Diseño de grupo control no equivalente con pretest-postest

El diseño de grupo control no equivalente con sólo postest puede mejorarse mucho con la aplicación de un pretest. Con esto, tenemos un **diseño de grupo control no equivalente con pretest-postest,** uno de los diseños cuasiexpe-rimentales más útiles. Se puede expresar de la siguiente manera:



No se trata de un diseño experimental completo, debido a que la asignación a los grupos no es aleatoria y a que es probable que los grupos no sean equivalentes. Sin embargo, usted tiene la ventaja de conocer los puntajes del pretest. De este modo podemos ver que los grupos eran similares en el pretest. Aun cuando los grupos no sean equivalentes, podemos observar *cambios* en los puntajes del pretest al postest. Si la variable independiente tiene algún efecto, el grupo experimental debe mostrar un mayor cambio que el grupo control (véase Kenny, 1979). Trochim (2000) explica las estrategias para el análisis estadístico de este cipo de cambio en los puntajes.

Joy, Kimball y Zabrack (1986) utilizaron un diseño de grupo control no equivalente con pretest-postest para estudiar el efecto de la televisión sobre la conducta agresiva de los niños. Un pueblo canadiense que no tuvo recepción televisiva sino hasta 1974 fue el centro de su estudio (Joy y sus colaboradores

bautizaron al pueblo como "Notel"). Antes y después de la introducción de la televisión en Notel, los investigadores midieron la agresión física y verbal de un grupo de niños. Al mismo tiempo, midieron la agresión en dos pueblos similares, uno de los cuales recibía únicamente un canal canadiense ("Unitel"), y otro que recibía cadenas de televisión canadienses y estadounidenses ("Multitel"). De esta manera, fue posible comparar los cambios en la agresión en Notel, con aquellos en las comunidades control de Unitel y Multitel. Los resultados del estudio mostraron que hubo un mayor incremento de la agresión en Notel que en Unitel y Multitel.

Diseño de series de tiempo interrumpidas

Campbell (1969) comenta extensamente la evaluación de una reforma legal específica: las medidas severas de 1955 en contra del exceso de velocidad en Co-nnecticut. A pesar de que se trata de un hecho muy antiguo, aún constituye un ejemplo adecuado de un aspecto metodológico importante. Las medidas severas se establecieron después de un número exageradamente alto de accidentes de tránsito fatales ocurridos en 1955. La forma más fácil de evaluar esta reforma es comparar la cantidad de accidentes de tránsito fatales en 1955 (antes de la aplicación de las medidas), con la cantidad en 1956 (después de la aplicación de las medidas). De hecho, hubo una reducción en la cantidad de muertes por accidentes de tránsito, de 324 en 1955 a 284 en 1956. Esta comparación constituye realmente un diseño de un grupo con pretest-postest, con todos los problemas de validez interna de este diseño; existen muchas otras razones por las que los accidentes de tránsito fatales pudieron haber disminuido. Una alternativa es el uso de un **diseño de series de tiempo interrumpidas,** que examinaría la tasa de accidentes mortales durante un largo periodo, antes y después de instituirse la reforma. La figura 11.2 presenta esta información para los años 1951-1959. Campbell (1969) argumenta que la disminución que hubo de 1955 a 1956 no es

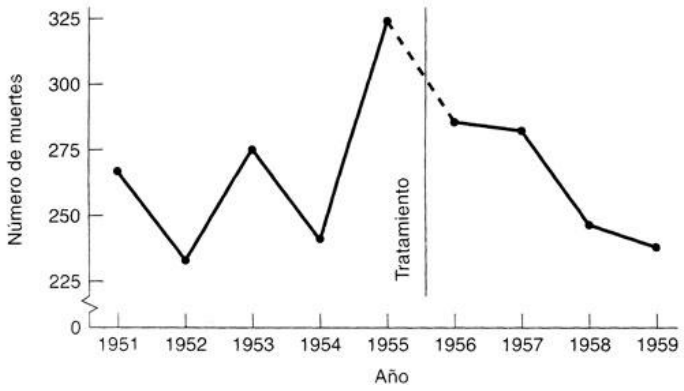


FIGURA 11.2
Accidentes de tránsito fatales en Connecticut, 1951-1959.

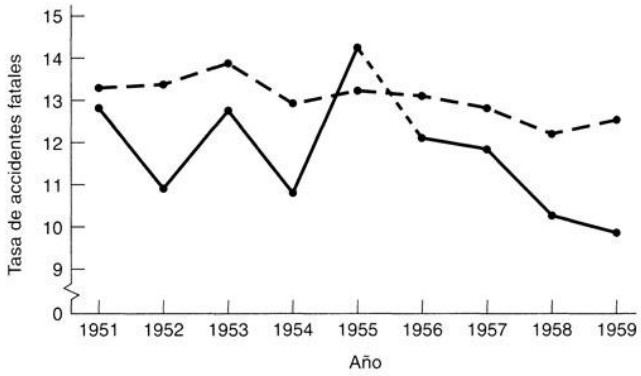


FIGURA 11.3
Diseño de series de control que comparan la tasa de accidentes de tránsito fatales en Connecticut (línea continúa) con la tasa de accidentes fatales de cuatro estados comparables (línea discontinua).

particularmente impresionante, dada la gran fluctuación en los años anteriores; aunque existe una tendencia decreciente estable en el número de muertes después de la reforma. Sin embargo, incluso en esto Campbell encuentra un problema de interpretación. La disminución puede deberse a la regresión estadística: puesto que 1955 fue un año con una tasa excesivamente alta, existe la probabilidad de que se diera una disminución de cualquier manera. Aun así, los datos de los años previos y posteriores a la reforma permiten una interpretación menos ambigua de la que podría hacerse únicamente con los datos de 1955 *y* 1956.

Diseño de series de control

Una manera de mejorar el diseño de series de tiempo interrumpidas es creando algún tipo de grupo control: un **diseño de series de control.** En la reforma del límite de velocidad de Connecticut fue posible este diseño debido a que otros estados no habían instituido la reforma. La figura 11.3 muestra los mismos datos sobre accidentes de tránsito fatales en Connecticut, más las cifras de accidentes fatales en cuatro estados comparables, durante los mismos años. El hecho de que las tasas de accidentes fatales en los estados control permanecieran relativamente constantes, mientras que las tasas de Connecticut disminuyeran de manera consistente, dio lugar a que Campbell concluyera que la reforma tuvo algún efecto.

Quizás se pregunte acerca de la evaluación del programa DARE. De hecho, muchos investigadores han llevado a cabo estudios de evaluación de resultados por medio de diseños cuasiexperimentales para examinar sus efectos a corto y largo plazos. La mayoría de los estudios comparan estudiantes de escuelas que incluyen programas DARE, con estudiantes de escuelas que no lo hacen. La conclusión general es que el DARE tiene efectos muy pequeños en los participantes (por ejemplo, Ennett, Tobler, Ringwalt y Flewelling, 1994). Asimismo, estudios que han examinado sus fectos a largo plazo concluyen que el programa no tiene efectos benéficos (Rosenbaum y Hanson, 1998); por ejemplo, estudiantes univer-

sitarios que participaron en el DARE cuando eran niños o adolescentes, mantienen el mismo consumo de sustancias que estudiantes que nunca fueron incluidos en el programa (Thombs, 2000).

DISEÑOS EXPERIMENTALES DE CASO ÚNICO

Los **diseños experimentales de caso único** se conocen tradicionalmente como diseños *de sujeto único,* aunque ahora se utilizan los términos *caso único* (Barlow y Hersen, 1984; Shadish *etal,* 2002) *y participante único* (Morgan y Morgan, 2001). Estos diseños tienen una larga historia de aplicación dentro del contexto de la investigación del refuerzo. Pueden encontrarse en el trabajo de B. F. Skinner (1953), sobre los programas de reforzamiento, y con frecuencia se utilizan en ambientes aplicados y clínicos, cuando se emplean técnicas de modificación conductual (Kazdin, 2001). Sin embargo, las técnicas y la lógica de los experimentos de caso único pueden aplicarse también a otras áreas de investigación.

Los experimentos de caso único se crearon a partir de la necesidad de determinar si una manipulación experimental tenía algún efecto sobre un participante de investigación. En el diseño de caso único el comportamiento del sujeto se mide a lo largo del tiempo, durante un periodo control de **línea base.** Después se introduce la manipulación durante el periodo de tratamiento y se sigue observando la conducta del sujeto. Un cambio en su comportamiento, del periodo de línea base al periodo del tratamiento, es evidencia de la eficacia de la manipulación. El problema, sin embargo, es que podrían existir muchas explicaciones para el cambio, distintas al tratamiento experimental (es decir, explicaciones alternas). Por ejemplo, es probable que otro suceso haya coincidido con la introducción del tratamiento. Los diseños de caso único descritos en las siguientes secciones abordan este problema.

Diseños reversibles

Como se señaló, el aspecto básico de los experimentos de caso único es la determinación de que la manipulación de la variable independiente tuvo un efecto. Un método es demostrar la reversibilidad de la manipulación. Un **diseño reversible** simple toma la siguiente forma:

$$\text{A (periodo de línea base)} \longrightarrow \text{B (periodo del tratamiento)} \longrightarrow \text{A (periodo de línea base)}$$

Este diseño, llamado ABA, requiere que se observe la conducta durante el periodo control de línea base (A), nuevamente durante el periodo de tratamiento (B) y también durante un segundo periodo de lineábase (A), una vez que el tratamiento experimental ha sido retirado (en ocasiones se *conoce* como *diseño de retiro,* por el hecho de que el tratamiento se retira o quita). Por ejemplo, el efecto de un procedimiento de reforzamiento sobre el desempeño académico

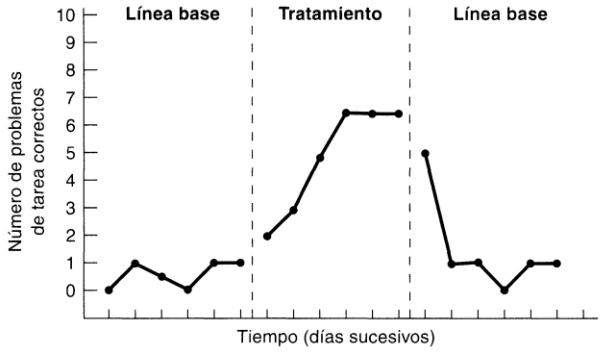


FIGURA 11.4
Datos hipotéticos de un diseño reversible ABA.

de un niño puede evaluarse por medio de un diseño ABA. Durante la línea base, se podría medir diariamente la cantidad de problemas de la tarea que resuelve de manera correcta. Después se introduciría un tratamiento con refuerzo, en el que el niño reciba estrellas por los problemas correctos, las cuales podría acumular e intercambiar por juguetes o dulces. Posteriormente, el tratamiento sería retirado durante el segundo periodo de línea base (A). En la figura 11.4 se muestran datos hipotéticos de un experimento como éste. El hecho de que la conducta cambie cuando se introduce el tratamiento y que regrese a su estado anterior al retirarlo, es evidencia de su eficacia.

La figura 11.4 describe un tratamiento que tuvo una repercusión relativamente dramática sobre el comportamiento. Algunos tratamientos producen un cambio inmediato en la conducta, aunque muchas otras variables podrían requerir de mayor tiempo para mostrar un efecto. Dermer y Hoch (1999) señalan que los diseños de caso único son adecuados para estas variables, al mantener el tratamiento durante un largo periodo.

El diseño ABA puede mejorarse de forma importante si se extiende a un diseño ABAB, en el que el tratamiento experimental se introduce una segunda vez, e incluso a un diseño ABABAB, el cual permite que los efectos del tratamiento sean probados en una tercera ocasión. Esto se hace para resolver dos problemas del diseño reversible ABA. Primero, una sola reversión no constituye evidencia sumamente poderosa con respecto de la eficacia del tratamiento. La reversión observada podría deberse a una fluctuación aleatoria de la conducta del niño, quizás el tratamiento coincidió con algún otro acontecimiento, como el próximo cumpleaños del niño, que provocó el cambio (y la reversión posterior al cumpleaños). Las posibilidades son mucho menores *si* se ha probado en dos o más ocasiones que el tratamiento muestra un efecto; es poco probable que sucesos aleatorios o coincidentes sean responsables de ambas reversiones. El secundo problema es de tipo ético. Tal como señalan Barlow y Hersen (1984), no es correcto finalizar el diseño con el retiro de un tratamiento que podría ser muy benéfico para el participante. El uso de un diseño ABAB brinda la oportu-

nidad de observar una segunda reversión cuando el tratamiento se introduce nuevamente. La secuencia termina en el tratamiento y no en su retiro.

En un diseño reversible también puede emplearse un grupo control. Por ejemplo, un equipo de investigadores utilizó el reforzamiento para incrementar el número de usuarios de un sistema de autobuses de un *campus* universitario (Everett, Hayward y Meyers, 1974). Se contó el número de usuarios en dos autobuses distintos durante 36 días. La manipulación experimental consistió en dar una ficha a cada usuario en uno de los autobuses marcados especialmente; las fichas podían intercambiarse por bienes y servicios en tiendas locales. En el autobús control no se utilizó fichas. Se empleó un diseño ABA. Los primeros 16 días constituyeron el periodo de línea base, y durante esta fase el número de usuarios diarios de cada autobús fue de alrededor de 250. La manipulación experimental se introdujo durante los días 17-24. Durante este periodo, el número de usuarios en el autobús experimental fue de aproximadamente 400 por día; en el autobús control no hubo cambio en el número de usuarios. El sistema de fichas se retiró el día 25 y se evaluó el número de usuarios de ambos autobuses hasta el día 36. La cantidad de usuarios del autobús experimental regresó a los niveles de la línea base durante este periodo, y ya no fue mayor que el del autobús control.

Diseños de línea base múltiple

Quizás a usted se le habrá ocurrido que la reversión de ciertas conductas es imposible o poco ética. Por ejemplo, sería poco ético revertir un tratamiento que reduce conductas peligrosas o ilegales, como el exhibicionismo o el alcoholismo; incluso, si existe la posibilidad de que una segunda introducción del tratamiento pueda provocar otro cambio. Otros tratamientos pueden producir un cambio de conducta a largo plazo que no sea reversible. En estos casos, se pueden realizar medidas múltiples a lo largo del tiempo, antes y después de la manipulación. Si la manipulación es eficaz, se observará de inmediato un cambio en la conducta, el cual continuará reflejándose en medidas posteriores. En un **diseño de línea base múltiple** la eficacia del tratamiento se demuestra cuando un comportamiento cambia, únicamente después de introducir la manipulación. Para demostrar la eficacia del tratamiento, dicho cambio deberá observarse bajo *múltiples* circunstancias para eliminar la posibilidad de que otros sucesos sean los responsables.

Existen variantes del diseño de línea base múltiple (Barlow y Hersen, 1984). En la línea base múltiple *entre sujetos* se mide el comportamiento de varios sujetos a lo largo del tiempo; sin embargo, para cada sujeto la manipulación se introduce en diferentes momentos. La figura 11.5 muestra datos de un experimento hipotético sobre reducción de tabaquismo con tres sujetos. Observe que la introducción de la manipulación fue seguida por un cambio en la conducta de cada uno de los sujetos. No obstante, puesto que este cambio ocurrió entre los individuos y la manipulación se introdujo en distintos momentos a cada sujeto, podemos descartar explicaciones basadas en el azar, eventos históricos, etcétera.

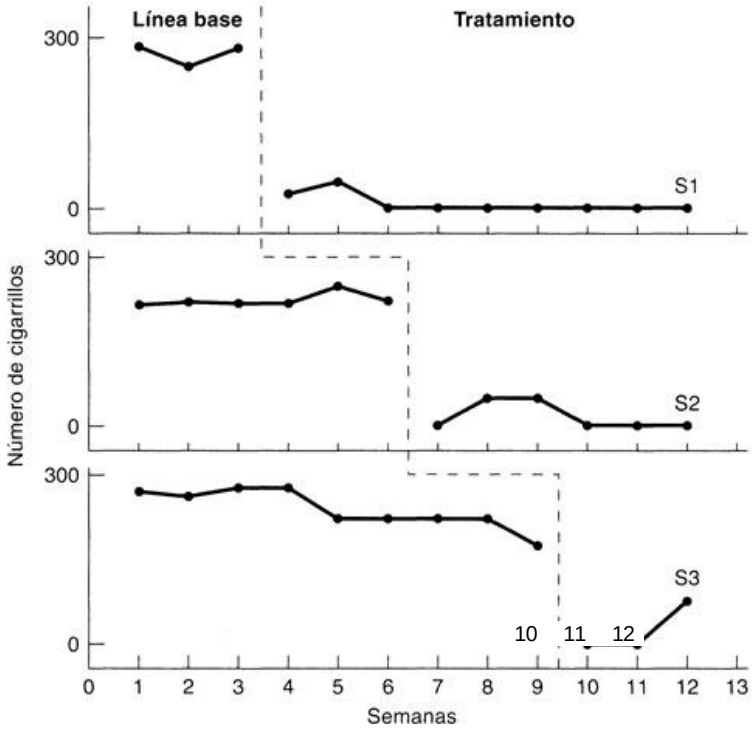


FIGURA 11.5
Datos hipotéticos de un diseño de línea base múltiple entre sujetos.

En una línea base múltiple *entre conductas*, se miden distintas conductas de un solo sujeto a lo largo del tiempo. En diferentes ocasiones se aplica la misma manipulación a cada una de las conductas. Por ejemplo, se puede establecer un sistema de recompensas para incrementar las conductas de socialización, aseo y lectura de un paciente psiquiátrico. El sistema de recompensas se aplicaría a cada uno de estos comportamientos en momentos distintos. La demostración de que cada conducta se incrementó al aplicar el sistema de recompensas sería una evidencia de la eficacia de la manipulación.

La tercera variante es la línea base múltiple *entre situaciones*, en la que se mide la misma conducta en distintos ambientes, como el hogar y el trabajo. Nuevamente, se introduce la manipulación en diferentes momentos en cada ambiente, con la esperanza de que ocurra un cambio de la conducta en cada situación, únicamente después de la manipulación.

Replicación de los diseños de caso único

Los procedimientos que se utilizan con un sujeto único pueden, por supuesto, reaplicarse con otros sujetos, lo cual aumenta enormemente la posibilidad de generalizar los resultados. Generalmente, los reportes de investigaciones que emplean procedimientos experimentales de caso único presentan los resultados

de varios sujetos (y con frecuencia en varios ambientes). La tradición en la investigación de caso único ha sido presentar los resultados de cada sujeto de manera individual y no como datos grupales, e incluir medias totales. Sidman (1960), vocero de esta tradición, ha señalado que agrupar los datos de una cantidad de sujetos, a través de medias grupales, en ocasiones puede dar una imagen confusa de las respuestas individuales a la manipulación. Por ejemplo, la manipulación puede ser eficaz para cambiar la conducta de algunos sujetos pero no de otros. Esto fue lo que sucedió en un estudio del uso de los cinturones de seguridad (Berry y Geller, 1991). Se analizaron distintas condiciones de señales de los cinturones de seguridad (por ejemplo, una segunda señal que aparecía si el sujeto no utilizaba el cinturón después de una señal inicial). De 13 sujetos, seis utilizaron siempre el cinturón, sin importar la condición, y tres nunca lo emplearon. Para los cuatro sujetos restantes, las condiciones con la señal sí tuvieron un efecto. Puesto que el énfasis del estudio eran los sujetos individuales, el patrón de resultados se reveló con rapidez.

Los diseños de caso único son útiles para estudiar muchos problemas de investigación y deben ser considerados como una alternativa poderosa a los diseños de investigación más tradicionales. Pueden ser especialmente valiosos para alguien que está aplicando alguna técnica de cambio en un ambiente natural; por ejemplo, un maestro que aplica una técnica nueva en el salón de clases. Además, no se requieren análisis estadísticos complejos para los diseños de caso único.

Hasta ahora hemos descrito la mayoría de las aproximaciones más importantes para diseñar investigaciones. En los siguientes dos capítulos estudiaremos métodos para analizar datos de investigación.

Términos de estudio

Diseño de grupo control no equivalente
Diseño de grupo control no equivalente con pretest-postest
Diseño de línea base múltiple
Diseño de series de control
Diseño de series de tiempo interrumpidas
Diseño de un grupo con pretest-postest
Diseño de un grupo con únicamente postest
Diseño reversible
Efectos de la historia
Efectos de maduración
Efectos de prueba
Evaluación de la eficacia
Evaluación de la teoría de programas
Evaluación de necesidades
Evaluación de programas
Evaluación de resultados
Evaluación del proceso
Experimento de caso único
Línea base
Problemas de instrumentación
Regresión estadística (regresión hacia la media)
Regresión hacia la media

Preguntas de repaso

1. Describa cuáles serían las metas de un investigador de evaluación de un programa, cuando aborda cada uno de los cinco tipos de preguntas de investigación de evaluación.
2. ¿Por qué un investigador utilizaría un diseño cuasiexperimental en lugar de un diseño experimental completo?
3. Describa las amenazas a la validez interna comentadas en el texto: la historia, la maduración, la prueba, los problemas de instrumentación, la regresión hacia la media y las diferencias de selección.
4. ¿Por qué tener un grupo control que elimina los problemas asociados con el diseño de un grupo con pretest-postest?
5. Describa el diseño de grupo control no equivalente con pretest-postest. ¿Por qué es éste un diseño cuasiexperimental y no un diseño experimental completo?
6. Describa los diseños de series de tiempo interrumpidas y de series de control. ¿Cuáles son las ventajas del diseño de series de control, en comparación con el diseño de series de tiempo interrumpidas?
7. ¿Qué es un diseño reversible? ¿Por qué un diseño ABAB es mejor que un diseño ABA?
8. ¿A qué se refiere la *línea base* en un diseño de caso único?
9. ¿Qué es un diseño de línea base múltiple? ¿Por qué se utiliza? Determine las diferencias entre los diseños de línea base múltiple entre sujetos, entre conductas y entre situaciones.

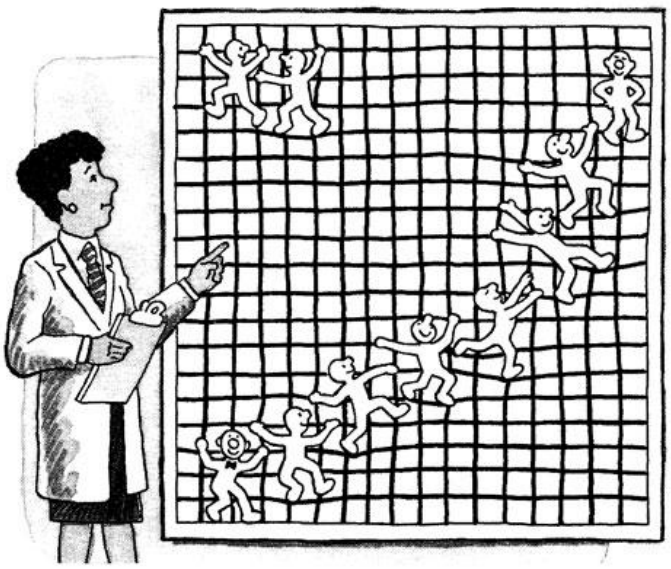
Actividades

1. El doctor Smith supo que una fraternidad de la universidad había comprado varias computadoras Macintosh, y que otra fraternidad había comprado varias computadoras con el sistema Windows. El doctor estaba interesado en saber si el tipo de computadora afecta la calidad de los trabajos de los estudiantes, por lo que acudió con cada una de las fraternidades para reunir muestras de trabajos de sus miembros. Después, dos estudiantes graduados, del departamento de inglés, calificaron la calidad de los trabajos. El doctor encontró que la calidad de los trabajos era mayor en una fraternidad que en la otra. ¿Cuáles son las variables independiente y dependiente en este estudio? Identifique el tipo de diseño que utilizó el doctor. ¿Qué variables se confunden con la variable independiente? Diseñe un experimento completo que responda la pregunta original del doctor Smith.

2. Gilovich (1991) describió un incidente que leyó durante una visita a Israel. Durante un corto periodo había ocurrido un gran número de muertes en una región del país. Un grupo de rabinos atribuyó las muertes a un cambio reciente en las prácticas religiosas, que permitía a las mujeres acudir a los funerales. De inmediato se prohibió a las mujeres asistir a los funerales en esta región, y el número de muertes disminuyó subsecuentemente. ¿Cómo podría explicar este fenómeno?
3. Los capitanes de cada zona de un departamento de policía metropolitana seleccionaron a dos oficiales para participar en un programa diseñado para reducir los prejuicios, al incrementar la sensibilidad ante las diferencias entre grupos raciales y étnicos y ante temas comunitarios. El programa de entrenamiento se llevó a cabo cada viernes por la mañana, durante tres meses. Durante la primera y la última sesión, los oficiales respondieron a una prueba de prejuicios. Para evaluar la eficacia del programa, se comparó el puntaje promedio de los prejuicios de la primera sesión, con el puntaje promedio de la última; se encontró que el puntaje promedio fue menor al finalizar el tratamiento. ¿De qué tipo de diseño se trata? ¿Qué problemas específicos surgen si se trata de concluir que el programa de entrenamiento fue responsable de la reducción de los prejuicios?
4. Un club de estudiantes está tratando de decidir si pone en marcha un programa de tutoría entre compañeros, para estudiantes que están en un curso de estadística de su departamento. Los miembros del club que ya tomaron el curso de estadística se ofrecerían como tutores de los estudiantes que actualmente están tomando el curso. Usted decide tomar en serio las ventajas de una investigación de evaluación de programas, por lo que diseña una estrategia para llevar a cabo un estudio de evaluación.
   a) ¿De qué forma mediría la necesidad de un programa como éste?
   b) Describa brevemente cómo podría aplicar un programa de tutoría. ¿Cómo verificaría el programa?
   c) Proponga un diseño cuasiexperimental para evaluar la eficacia del programa.
   d) ¿Cómo determinaría la eficiencia económica de un programa como éste?
5. Muchas escuelas primarias han implantado un periodo diario de "lectura en silencio sostenida", durante el cual los estudiantes, los maestros y el personal pasan de 15 a 20 minutos leyendo en silencio un libro de su elección. Los partidarios de esta política afirman que la actividad motiva la lectura placentera fuera del tiempo de lectura silenciosa requerido. Diseñe un cuasiexperimento con grupo control no equivalente, con pretest-postest, para probar esta afirmación. Incluya también una medida dependiente pensada adecuadamente.

6. Para la situación anterior, discuta las ventajas y desventajas del uso de un diseño cuasiexperimental, en contraste con la realización de un experimento completo.
7. Su perro se siente solo mientras usted está en el trabajo y, como consecuencia, realiza actividades destructivas como jalar las cortinas o vaciar el contenido del basurero en el piso. Usted decide que quizás sería útil encender el radio mientras está ausente. ¿Cómo podría determinar si este "tratamiento" es eficaz?
8. Su mejor amigo con frecuencia padece fuertes dolores de cabeza. Usted ha notado que su amigo consume grandes cantidades de refresco de cola dietético, por lo cual considera la hipótesis de que el endulzante artificial del refresco es el responsable de los dolores de cabeza. Diseñe una forma para probar la hipótesis por medio de un diseño de caso único. ¿Qué espera encontrar si su hipótesis es correcta? Si obtienen los resultados esperados, ¿qué concluye acerca del efecto que el endulzante artificial tiene sobre los dolores de cabeza?

12



Comprensión de los resultados de la investigación: descripción y correlación

Escalas de medición: repaso

Análisis de los resultados de la investigación
Comparación de porcentajes grupales Correlación de puntajes individuales Comparación de medias grupales

Distribuciones de frecuencias
Gráficas de distribuciones de frecuencias *Gráficas de pastel Gráficas de barras Polígonos de frecuencias*

Estadística descriptiva
Tendencia central Variabilidad

Gráficas de relaciones

Coeficientes de correlación: descripción de la fuerza de relaciones
Coeficiente de correlación r de Pearson
Consideraciones importantes
Restricción del rango
Relación curvilínea

Magnitud del efecto

Significación estadística

Ecuaciones de regresión

Correlación múltiple

Correlación parcial y el problema de la tercera variable

Modelos estructurales

Términos de estudio

Preguntas de repaso

Actividades

La estadística nos ayuda a comprender los datos obtenidos en las investigaciones. Existen dos razones para emplear la estadística. Primero, se utiliza para describir los datos. En segundo lugar, se usa para hacer inferencias, con base en datos muéstrales, acerca de una población. En este capítulo estudiamos la estadística descriptiva y la correlación; en el capítulo 13 abordamos la estadística inferencial. Nos enfocaremos principalmente en la lógica subyacente y en los procedimientos generales para tomar decisiones estadísticas. En el apéndice B se incluyen los cálculos específicos de diversos análisis estadísticos.

ESCALAS DE MEDICIÓN: REPASO

Antes de abordar la estadística, necesitamos repasar el concepto de escalas de medición. Siempre que se estudia una variable, debe definirse operacionalmente e incluirse dos o más niveles de ésta. En el capítulo 5 estudiamos que los niveles de la variable pueden describirse por medio de una de cuatro escalas de medición: nominal, ordinal, de intervalo y de razón. La escala utilizada determina el tipo de estadística apropiada para el análisis de los resultados de un estudio. Recuerde que el significado de un puntaje específico en una variable depende del tipo de escala que se empleó cuando se midió o manipuló.

Los niveles de las variables en una escala nominal no tienen propiedades numéricas o cuantitativas. Los niveles únicamente representan categorías o grupos diferentes. La mayoría de las variables independientes en los experimentos son nominales, como sucede en un experimento que compara una terapia conductual con una terapia cognoscitiva para la depresión. Las variables como el género, el color de los ojos, la mano dominante, el orden de nacimiento y el estado civil son variables nominales; los individuos zurdos y diestros son simplemente distintos.

Los niveles de las variables con escala ordinal implican diferencias cualitativas mínimas. Es posible ordenar los niveles de la variable, del más bajo al más alto. El ejemplo más claro de una escala ordinal es pedir a las personas que hagan juicios de orden. Por ejemplo, se les puede pedir que ordenen los problemas más importantes que enfrenta el estado donde viven. Si en primer lugar está la educación, la salud en segundo y el crimen en tercero, conocemos el orden; pero no sabemos cómo se sienten las personas por estos problemas. La educación y la salud podrían aproximarse mucho en su importancia, y el crimen está en el tercer lugar más lejano; el intervalo entre cada uno de los problemas no es igual.

Las variables de intervalo y de razón poseen propiedades cuantitativas más detalladas. En una variable de intervalo, los intervalos entre los niveles son del mismo tamaño. La diferencia entre uno y dos en la escala, por ejemplo, es igual a la diferencia entre dos y tres. Generalmente, las escalas de intervalo incluyen cinco o más niveles cuantitativos. Usted puede pedir a la gente que califique su estado de ánimo en una escala de siete puntos, que va de "muy negativo" a "muy positivo". No existe un cero absoluto que indique la "ausencia" de estado de animo.

En las ciencias del comportamiento suele ser difícil saber con precisión si se está utilizando una escala ordinal o una de intervalo. Sin embargo, es útil suponer que la variable se está midiendo en una escala de intervalo, ya que las escalas de intervalo permiten tratamientos estadísticos más sofisticados que las escalas ordinales. Por supuesto, si la medida es un orden de rangos (por ejemplo, estudiantes en una clase ordenados con base en su popularidad), obviamente se trata de una escala ordinal.

La escala de razón incluye intervalos iguales y un cero absoluto que indica la ausencia de la variable que se está midiendo. El tiempo, el peso, la longitud y otras medidas físicas constituyen los mejores ejemplos de escalas de razón. Las variables de intervalo y de razón son conceptualmente distintas; no obstante, los procedimientos estadísticos que se utilizan para analizar los datos de este tipo de variables son idénticos. Una implicación importante de las escalas de intervalo y de razón es que los datos pueden resumirse utilizando la media o promedio aritmético. Se puede obtener una cifra que refleje la cantidad media de una variable; por ejemplo, el "estado de ánimo promedio de las personas que ganaron un concurso fue de 5.1" o el "peso medio de los hombres que completaron el programa de pérdida de peso fue de 187.7".

ANÁLISIS DE LOS RESULTADOS DE LA INVESTIGACIÓN

Las escalas de medición tienen implicaciones importantes en la forma en que se describen y analizan los resultados de las investigaciones. Éstas, en su mayoría, se centran en el estudio de relaciones entre variables. Según la forma en que se analizan las variables, hay tres formas básicas para describir los resultados: 1) comparación de porcentajes grupales, 2) correlación de puntajes de individuos en dos variables y 3) comparación de medias grupales.

Comparación de porcentajes grupales

Suponga que desea saber si los hombres y las mujeres difieren en su interés por viajar. En su estudio, usted pregunta a hombres y mujeres si les gusta o no viajar. Para describir sus resultados, necesita calcular el porcentaje de mujeres a las que les gusta viajar y compararlo con el porcentaje de hombres que también disfrutan de esta actividad. Suponga que probó 50 mujeres y 50 hombres y que encontró que a 40 mujeres y a 30 hombres les gusta viajar. Para describir sus hallazgos, reportaría que a 80 por ciento de las mujeres les gusta viajar, en comparación con el 60 por ciento de los hombres. Por lo tanto, aparentemente existe una relación entre las variables género y gusto por viajar. Observe que nos enfocamos en porcentajes debido a que el gusto por viajar es una variable nominal: el gusto y el disgusto son simplemente dos categorías distintas. Después de describir sus datos, el siguiente paso es la realización de un análisis estadístico para determinar si existe una diferencia estadísticamente

significativa entre los hombres y las mujeres. La significación estadística se estudia en el capítulo 13; los procedimientos de análisis estadísticos se describen en el apéndice B.

Correlación de puntajes individuales

Cuando no hay distintos grupos de sujetos, se requiere de un segundo tipo de análisis. Los individuos se miden en dos variables, y cada variable tiene un rango de valores numéricos. Más adelante, en este capítulo, abordaremos un análisis de datos sobre la relación entre el asiento que se ocupa en el salón de clases y las calificaciones: ¿las personas que se sientan al frente reciben calificaciones más altas?

Comparación de medias grupales

Numerosas investigaciones se diseñan para comparar las medias de las respuestas de los participantes de dos o más grupos. Por ejemplo, en un experimento diseñado para estudiar los efectos de verse expuesto a un adulto agresivo, los niños de un grupo observan a un adulto "modelo" que se comporta de manera agresiva, mientras que los niños de un grupo control no. Después, cada niño juega solo durante diez minutos en una habitación que contiene cierta cantidad de juguetes, y un grupo de observadores registra el número de veces que el niño se comporta de manera agresiva durante el juego. El número de agresiones es una variable de razón, ya que en la escala existen intervalos iguales y un cero absoluto.

En este caso, nos interesaríamos en comparar el número promedio de actos agresivos de los niños en ambas condiciones para determinar si aquellos que observaron al modelo fueron más agresivos que los niños de la condición control. En la tabla 12.1 se muestran datos hipotéticos de este experimento, donde se incluyen diez niños en cada condición; los puntajes de la tabla representan el número de actos violentos de cada niño. En este caso, el puntaje promedio de la agresividad del grupo con modelo es de 5.20 y la del grupo sin modelo es de 3.10. En el próximo capítulo realizaremos una prueba estadística para determinar si esta diferencia es estadísticamente significativa.

Para cualquier tipo de datos, es importante comprender los resultados describiendo cuidadosamente la información reunida. Iniciamos con la construcción de la distribución de frecuencias.

DISTRIBUCIONES DE FRECUENCIAS

Cuando se analizan los resultados, es útil iniciar con la elaboración de una distribución de frecuencias de los datos. Una **distribución de frecuencias** indica la cantidad de individuos que reciben cada puntaje posible de una variable. Las distribuciones de frecuencias de calificaciones de exámenes son familiares para

TABLA 12.1 Puntajes de una medida de agresión en un experimento hipotético sobre el modelamiento de la agresión.

Grupo con modelo	Grupo sin modelo
3	1
4	2
5	2
5	3
5	3
5	3
6	4
6	4
6	4
7	5
$\Sigma X = 52$	$\Sigma X = 31$
$\bar{X} = 5.20$	$\bar{X} = 3.10$
$s^2 = 1.289$	$s^2 = 1.433$
$s = 1.135$	$s = 1.197$
$n = 10$	$n = 10$

la mayoría de los estudiantes universitarios; indican cuántos estudiantes recibieron una calificación dada en el examen. Aunado al número de individuos asociados a cada respuesta o puntaje, es útil examinar el porcentaje relacionado con esta cifra.

Gráficas de distribuciones de frecuencias

Es útil hacer una representación gráfica de las distribuciones de frecuencias. Veamos varios tipos de gráficas: de pastel, de barras y el polígono de frecuencias.

Gráficas de pastel Éstas dividen un círculo completo o "pastel" en "rebanadas" que representan porcentajes relativos. La figura 12.1 presenta una gráfica de pastel (en este caso, tridimensional) que muestra una distribución de frecuencias en la que a 70 por ciento de las personas les gusta viajar y al 30 por ciento no. Puesto que hay dos tipos de información en la gráfica, existen dos rebanadas en este pastel. Estas gráficas son especialmente útiles para representar información de escalas nominales. En la figura, el número de personas que eligió cada respuesta fue convertido a un porcentaje; por supuesto, se puede presentar únicamente el número.

Gráficas de barras En éstas se utiliza una barra separada y distinta para cada pedazo de información. La figura 12.2 representa la misma información sobre el gusto por viajar, por medio de una gráfica de barras. En esta gráfica, el eje *x*, u

horizontal, muestra las dos respuestas posibles. El eje y, o vertical, muestra el número de personas que eligió cada respuesta y así la altura de cada barra representa la cantidad de personas que respondieron a las opciones "gusto" y "disgusto".

Polígonos de frecuencias Estos utilizan una línea para representar frecuencias. Es un recurso muy útil cuando los datos corresponden a una escala de intervalo o de razón, como en los datos del modelamiento y la agresión presentados en la tabla 12.1. Aquí, claramente tenemos una escala numérica de la cantidad de actos agresivos ocurridos durante el periodo de observación. La figura 12.3 es una gráfica de los datos del experimento hipotético, que incluye dos polígonos de frecuencias, uno para cada grupo. La línea continua representa al grupo sin modelo y la punteada, al grupo con modelo.

FIGURA 12.1
Gráfica de pastel.

FIGURA 12.2
Gráfica de barras de datos obtenidos de dos grupos.

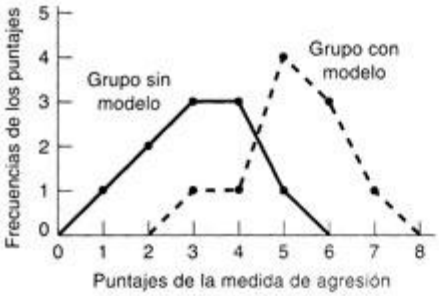


FIGURA 12.3
Polígonos de frecuencias que muestran las distribuciones de puntajes de la tabla 12.1.
Nota: cada polígono de frecuencias se origina y finaliza en puntajes que ningún niño obtuvo (0 y 6 en el grupo sin modelo, 2 y 8 en el grupo con modelo).

¿Qué puede descubrir cuando examina distribuciones de frecuencias? En primer lugar, puede observar de manera directa la forma en que respondieron los participantes. Puede ver cuáles son los puntajes más repetidos y su forma de distribución; puede descubrir la existencia de "puntajes extremos", poco comunes, inesperados o muy diferentes de los de otros participantes. En un experimento se puede comparar la distribución de los puntajes en los grupos.

ESTADÍSTICA DESCRIPTIVA

Además de examinar la distribución de los puntajes, usted puede calcular estadísticos descriptivos, los cuales permiten que los investigadores hagan afirmaciones precisas acerca de los datos. Para **describir los datos** se requieren dos estadísticos. Se puede utilizar un solo número para describir la tendencia central o los puntajes generales de los participantes. Otro número describe la variabilidad o qué tanto se dispersa la distribución de los puntajes. Estos dos números resumen la información contenida en una distribución de frecuencias.

Tendencia central

Una estadística de **tendencia central** nos indica cómo es, en promedio, la muestra considerada como un todo. Existen tres medidas de tendencia central: la media, la mediana y la moda. La **media** de un conjunto de puntajes se obtiene al sumarlos todos y dividir el resultado entre el número de puntajes; se simboliza con X, y en los reportes científicos se abrevia *M.* La media es un indicador de tendencia central apropiado únicamente cuando los puntajes se miden en una escala de intervalo o de razón, debido a que se utilizan los valores reales de los números para calcular la estadística. En la tabla 12.1, el puntaje de la media del grupo sin modelo es 3.10, y la del grupo con modelo es 5.20. Observe que la letra griega E (sigma) en la tabla 12.1 es la notación estadística para la sumatoria de un conjunto de números. Así, *YX* es la abreviatura de "la suma de los valores de un conjunto de puntajes".

La **mediana** es el puntaje que divide al grupo en mitades (con 50 por ciento de los puntajes por debajo y el otro 50 por ciento por arriba de la mediana). En los reportes científicos, la mediana se abrevia *Mdn {Md* en español). La mediana es apropiada cuando los puntajes están en una escala ordinal, ya que únicamente toma en cuenta el orden de los puntajes. Sin embargo, también es útil con variables de intervalo y de razón. La mediana del grupo sin modelo es tres y del grupo con modelo es *cinco.*

La **moda** es el puntaje más frecuente y la única medida de tendencia central que es adecuada para una escala nominal. No utiliza los valores reales de la escala, sino que solamente indica el valor que ocurre con mayor frecuencia. En el grupo sin modelo existen dos valores modales, el tres y el cuatro, que ocurren con la misma, frecuencia. La moda del grupo con modelo es cinco.

La mediana o la moda pueden ser mejores indicadores de tendencia central que la media, si algunos puntajes poco comunes sesgan la media. Por ejemplo, la mediana del ingreso familiar de un país o estado suele ser una mejor medida de tendencia central que la media de esta variable. Puesto que solamente una pequeña cantidad de individuos tiene ingresos extremadamente altos, el uso de la media haría parecer que la persona "promedio" gana más dinero del que en realidad obtiene.

Variabilidad

También podemos determinar cuánta **variabilidad** existe en un conjunto de puntajes. Una medida de variabilidad es un número que caracteriza la cantidad de dispersión en una distribución de puntajes. Una de estas medidas es la **desviación estándar,** que se simboliza con *s (d* en español), la cual indica el promedio de la desviación de los puntajes a partir de la media. En los reportes científicos se abrevia *SD (DE* en español). La desviación estándar se deriva calculando primero la **varianza,** cuyo símbolo es s^2 (d^2 en español). La desviación estándar es la raíz cuadrada de la varianza. La desviación estándar de un conjunto de puntajes es pequeña cuando la mayoría de las personas tienen puntajes similares y cercanos a la media. La desviación estándar se vuelve más grande conforme más personas tienen puntajes que caen lejos del valor de la media. Para el grupo con modelo, la desviación estándar es 1.135, la cual nos indica que la mayoría de los puntajes en dicha condición están a 1.135 unidades por arriba y por debajo de la media; es decir, entre 4.065 y 6.335. Por lo tanto, la media y la desviación estándar proporcionan una gran cantidad de información acerca de la distribución. Observe que, como sucede con la media, el cálculo de la desviación estándar utiliza los valores reales de los puntajes; en consecuencia, la desviación estándar únicamente es adecuada para variables de intervalo y de razón.

Otra medida de variabilidad es el *rango,* que simplemente es la diferencia entre el puntaje más alto y el más bajo. El rango de los grupos con modelo y sin modelo es de cuatro.

GRÁFICAS DE RELACIONES

La graficación de las relaciones entre variables se estudió brevemente en el capítulo 4. Una forma común de graficar las relaciones entre variables es el uso de una gráfica de barras o de una gráfica lineal. La figura 12.4 es una gráfica de barras que describe las medias de los grupos con modelo y sin modelo. Los niveles de la variable independiente (sin modelo y con modelo) están representadas en el eje *x* horizontal, y los valores de la variable dependiente se presentan en el eje *y* vertical. Para cada grupo, se escribe un punto sobre el eje *y,* que representa la media de los grupos, y luego se traza una barra para representar visualmente el valor medio. Las gráficas de barras se utilizan cuando los valores en el eje *x* son

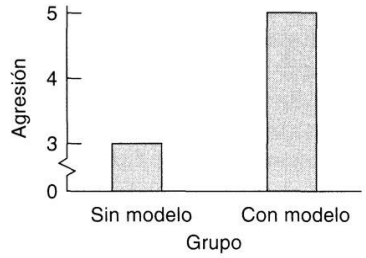


FIGURA 12.4
Gráfica de los resultados del experimento de modelamiento.

categorías nominales (por ejemplo, condiciones con y sin modelo). Las gráfica de líneas se usan cuando los valores en el eje *x* son numéricos) por ejemplo, la cantidad de horas que trabajan los adolescentes, como se muestra en la figura 7.1. En este caso, se hizo una línea para unir los puntos de los datos y representar la relación entre las variables.

Es interesante señalar un truco común que en ocasiones emplean los cienti-ficos y es muy utilizado por los anunciantes. Se trata de exagerar la distancia entre los puntos en la escala de medición para hacer que los resultados parezcan más impresionantes de lo que realmente son. Suponga, por ejemplo, que una empresa de refrescos de cola (cola A) lleva a cabo una prueba de sabor que indica que 52 por ciento de los participantes prefieren el refresco de cola A, y que 48 por ciento prefieren el refresco de cola B. ¿Cómo debe presentar la empresa los resultados? Las dos gráficas de barras de la figura 12.5 muestran el método más honesto, así como uno que es mucho más impresionante. Siempre es bueno observar cuidadosamente las cifras en las escalas descritas en las gráficas.

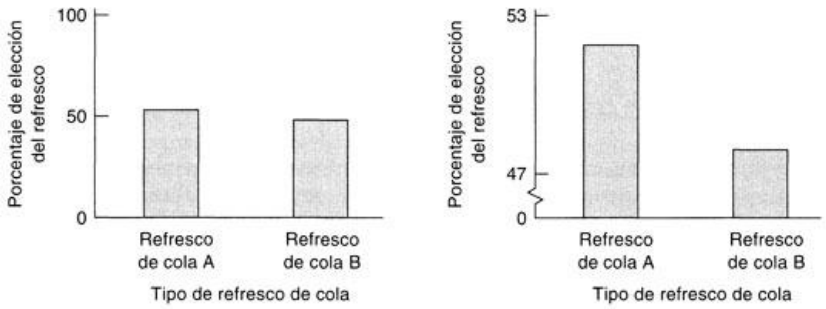


FIGURA 12.5
Dos formas de graficar los mismos datos.

COEFICIENTES DE CORRELACIÓN: DESCRIPCIÓN DE LA FUERZA DE RELACIONES

Es importante saber si la relación entre variables es relativamente débil o fuerte. Un **coeficiente de correlación** es un estadístico que describe la fuerza con la que están relacionadas las variables. Quizás usted esté más familiarizado con el **coeficiente de correlación productor/momento de Pearson,** el cual se utiliza cuando ambas variables tienen propiedades de escalas de intervalo o de razón. El coeficiente de correlación productor/momento de Pearson se denomina *r* de Pearson; sus valores pueden ser desde 0.00 hasta ± 1.00. Por lo tanto, la *r* de Pearson proporciona información acerca de la fuerza y dirección de la relación. Una correlación de 0.00 indica que no existe relación entre las variables. Mientras más se acerca una correlación a 1.00 (positivo o negativo), más fuerte es la relación. De hecho, a una correlación de 1.00 en ocasiones se le denomina relación perfecta, debido a que las variables varían de manera conjunta, en un patrón perfecto. El signo de la r de Pearson nos habla sobre la dirección de la relación; es decir, *si* existe una relación positiva o negativa entre las variables.

Los datos de estudios que examinan las similitudes que hay en los puntajes de pruebas de inteligencia entre hermanos, ilustran la conexión entre la magnitud del coeficiente de correlación y la fuerza de una relación. La relación que existe entre los puntajes de gemelos idénticos es muy fuerte (correlación de 0.86), lo cual demuestra una gran similitud entre los puntajes de las pruebas en estos pares de individuos. La correlación entre los gemelos fraternos criados juntos es menos fuerte, de 0.60. La correlación que existe entre hermanos que no son gemelos, que crecen juntos, es de 0.47, y la correlación entre hermanos que no son gemelos, criados de forma separada, es de 0.24 (véase Bouchard y McGue, 1981).

Existen muchos tipos diferentes de coeficientes de correlación. Cada coeficiente se calcula de una manera distinta, dependiendo del tipo de escala de medición de las dos variables. Como se señaló antes, el coeficiente de correlación r de Pearson es adecuado cuando los valores de ambas variables son de intervalo o de razón. Ahora nos concentraremos en los detalles del coeficiente de correlación producto-momento de Pearson.

Coeficiente de correlación r *de Pearson*

Para calcular un coeficiente de correlación necesitamos obtener pares de observaciones de cada sujeto. Así, cada individuo tiene dos puntajes, uno de cada variable. La tabla 12.2 muestra datos ficticios de diez estudiantes, con mediciones en las variables del asiento que ocupan en el salón de clases y la calificación de un examen. Los estudiantes de la primera fila reciben un puntaje de uno, los de la segunda fila un puntaje de dos, y así sucesivamente. Una vez hechas las observaciones, podemos ver si las dos variables están relacionadas. ¿Están relacionadas las variables de forma sistemática?

TABLA 12.2 Pares de puntajes de diez participantes del asiento que ocupan en clase y de las calificaciones de un examen (datos ficticios).

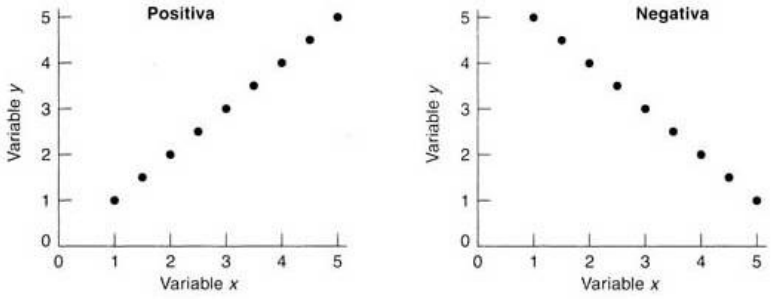
Número de identificación del sujeto	Asiento	Calificación del examen
01	2	95
02	5	50
03	1	85
04	4	75
05	3	75
06	5	60
07	2	80
08	3	70
09	1	90
10	4	70

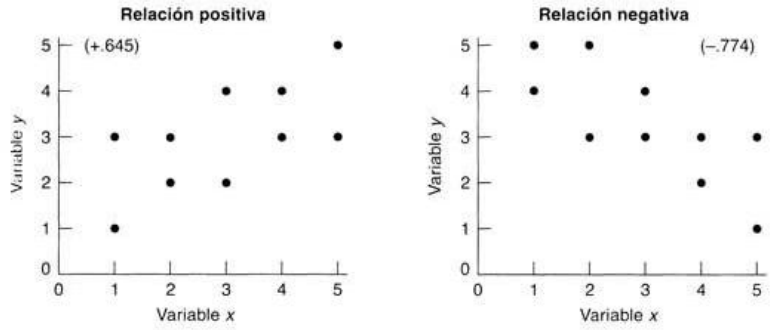
La *r* de Pearson brinda dos tipos de información acerca de la relación entre las variables. El primero es la fuerza de la relación, el segundo es la dirección de la relación. Como se señaló anteriormente, los valores de r pueden oscilar desde 0.00 hasta ± 1.00. El tamaño absoluto de r es el coeficiente que indica la fuerza de la relación. Un valor de 0.00 indica que no existe una relación. Entre más se acerca el valor de la r a 1.00 (positivo o negativo), más fuerte es la relación. Los signos más y menos indican la existencia de una relación lineal positiva o lineal negativa entre las dos variables. Es importante recordar que es el tamaño del coeficiente de correlación, y no el signo, el que indica la fuerza de la relación. Pollo tanto, un coeficiente de correlación de -0.54 indica una relación más fuerte que un coeficiente de + 0.45.

Los datos de la tabla 12.2 pueden visualizarse en un *diagrama de dispersión,* en el que cada par de puntajes se dibuja como un solo punto. La figura 12.6 muestra dos diagramas de dispersión. Los valores de la primera variable se encuentran en el eje *x, y* los de la segunda variable en el eje *y.* El diagrama de dispersión muestra una relación positiva perfecta (+1.00) y una relación negativa perfecta (-1.00). Es fácil ver por qué son relaciones perfectas: los puntajes en las dos variables forman una línea recta sobre la diagonal del diagrama. El puntaje de cada persona en una variable se ajusta perfectamente con su puntaje en la otra variable. Si conocemos el puntaje de un individuo en una de las variables, podemos decir exactamente cuál será su puntaje en la otra variable. Este tipo de relaciones "perfectas" se observan pocas veces en la realidad.

Los diagramas de dispersión de la figura 12.7 muestran patrones de correlación que tienen mayores posibilidades de encontrarse al explorar hallazgos de investigaciones. El primer diagrama presenta pares de puntajes con una correlación positiva de + 0.65; el segundo presenta una relación negativa de -0.77. Los puntos de estos dos diagramas de dispersión revelan un patrón general de una

FIGURA 12.6
Diagramas de dispersión de relaciones perfectas (± 1.00).





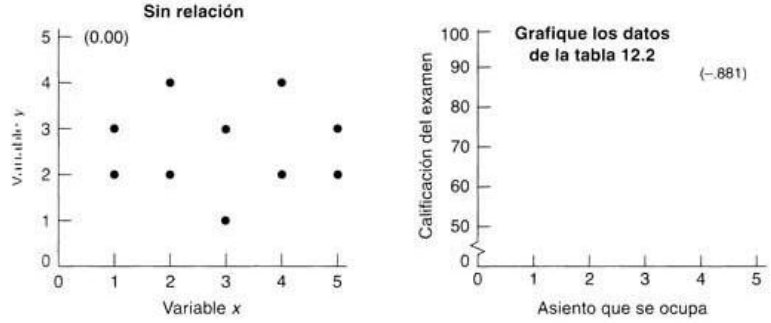


FIGURA 12.7
Diagramas de dispersión de patrones de correlación.

relación positiva o negativa, pero las relaciones no son perfectas. En el primer diagrama se puede hacer la predicción general, por ejemplo, de que a mayor puntaje en una variable, mayor puntaje en la segunda. Sin embargo, si usted *conoce* el puntaje de una persona en la primera variable, no puede predecir con exactitud cuál será su puntaje en la segunda. Para confirmar esto, observe el valor 1 en la variable *x* (el eje horizontal), en el diagrama de dispersión positivo. Verá que dos individuos tienen un puntaje de uno. Alguno de ellos tuvo un puntaje de uno en la variable *y* (el eje vertical) y el otro tuvo un puntaje de tres. Los datos no forman una diagonal perfecta, sino que existe una variación (dispersión) a partir de una línea diagonal perfecta.

En el tercer diagrama de dispersión no existe ninguna correlación ($r = 0.00$). Los puntos se distribuyen en todo el diagrama, con un patrón totalmente aleatorio. Por lo tanto, los puntajes de la variable *x* no están relacionados con los puntajes de la variable *y*.

El cuarto diagrama se dejó en blanco para que usted grafique los puntajes de los datos de la tabla 12.2. El eje *x* (horizontal) representa la variable del asiento que se ocupa y el eje *y* (vertical), la variable de la calificación del examen. Para elaborar el diagrama de dispersión, es necesario que grafique los diez pares de puntajes. Para cada individuo de la muestra encuentre el puntaje de la variable del asiento que se ocupa; después vaya hacia arriba, hasta encontrar la calificación de la persona en el examen. Ese punto describe el puntaje de ambas variables. En el diagrama de dispersión final habrá diez puntos.

El coeficiente de correlación calculado a partir de estos datos muestra una relación negativa entre las variables ($r = -0.88$). En otras palabras, conforme aumenta la distancia del asiento con relación al frente de la clase, la calificación del examen disminuye. Aun cuando estos datos son ficticios, la relación negativa es consistente con los hallazgos de investigaciones reales (Brooks y Rebata, 1991).

Consideraciones importantes

Restricción del rango Es importante que el investigador tome su muestra del rango más completo posible de valores de ambas variables. Si el rango de valores posibles está restringido, se reduce la magnitud del coeficiente de correlación. Por ejemplo, si el rango de los puntajes del asiento que se ocupa en el salón de clases se restringe a las dos primeras filas, usted no obtendrá una imagen precisa de la relación entre el asiento y la calificación del examen. De hecho, si sólo se toman en cuenta los puntajes de los estudiantes que se sientan en las dos primeras filas, la correlación entre las dos variables es exactamente 0.00. Con un rango limitado, existe una variabilidad restringida en los puntajes *y*, por lo tanto, menor variabilidad que pueda ser explicada.

El problema de un rango restringido se presenta cuando los individuos en la muestra son muy similares u *homogéneos* en la variable que se está estudiando. Por ejemplo, si usted estudia la edad como variable, probar únicamente a niños de seis y siete años reduce las posibilidades de encontrar efectos de la edad. De la

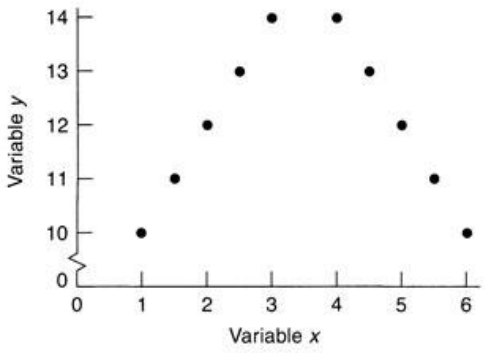


FIGURA 12.8
Diagrama de dispersión de una relación curvilínea (coeficiente de correlación producto-momento de Pearson = *0.00).*

misma forma, el estudio de las correlaciones de la inteligencia sería casi imposible *si* todos los miembros de la muestra tienen puntajes de inteligencia muy similares (por ejemplo, la clase del último año de una prestigiosa universidad privada).

Relación curvilínea El coeficiente de correlación producto-momento de Pearson (r) está diseñado para detectar únicamente relaciones lineales. Si la relación es curvilínea, como en el diagrama de dispersión de la figura 12.8, el coeficiente de correlación no indica la existencia de una relación. El coeficiente de correlación *r* de Pearson, calculado a partir de estos datos, es de exactamente 0.00, aun cuando las variables están claramente relacionadas.

Cuando la relación es curvilínea, se debe utilizar otro tipo de coeficiente de correlación para determinar la fuerza de la relación. Puesto que una relación puede ser curvilínea, es importante elaborar diagramas de dispersión, además de ver la magnitud del coeficiente de correlación. El diagrama de dispersión es valioso debido a que ofrece una imagen visual de la forma de la relación. Los programas de computadora que realizan análisis estadísticos generalmente muestran diagramas de dispersión y pueden indicarnos qué tan bien se ajustan los datos a una relación lineal o curvilínea.

MAGNITUD DEL EFECTO

Hemos descrito al coeficiente de correlación *r* de Pearson como la forma apropiada para mostrar la relación entre dos variables de intervalo o de razón. Los investigadores buscan poder describir la fuerza de las relaciones entre variables en todos los estudios. La **magnitud del efecto** es un término general que se refiere a la fuerza de la asociación entre variables. El coeficiente de correlación *r* de Pearson es un indicador de la magnitud del efecto, indica la fuerza de la asociación lineal entre dos variables. En un experimento con dos o más condiciones de tratamiento, se pueden calcular otros tipos de coeficientes de correlación para indicar la magnitud del efecto de la variable independiente sobre la varia-

ble dependiente. Por ejemplo, en nuestro experimento acerca de los efectos de observar a un modelo agresivo sobre la conducta agresiva de los niños, comparamos las medias de los dos grupos. Además de saber las medias, es útil conocer la magnitud del efecto. La magnitud del efecto calculada con un coeficiente de correlación puede ser ilustrada por medio del experimento de modelamiento y agresión. En este caso, el valor de la correlación de la magnitud del efecto es de 0.68. Como sucede con todos los coeficientes de correlación, los valores de esta correlación de la magnitud del efecto pueden oscilar entre 0.00 y 1.00 (no necesitamos preocuparnos por la dirección de la relación, por lo cual no se utilizan valores positivos o negativos). En el capítulo 13 se explica la fórmula utilizada para calcular la correlación.

La ventaja de reportar la magnitud del efecto es que nos proporciona una escala de valores que es consistente para todos los tipos de estudios. Los valores van desde 0.00 hasta 1.00, sin importar las variables empleadas, el diseño de investigación seleccionado ni la cantidad de participantes estudiados. Quizá se pregunte cuáles coeficientes de correlación deben considerarse indicadores de efectos pequeños, medianos o grandes. Una guía general es que las correlaciones cercanas a 0.15 (de alrededor de 0.10 a 0.20) se consideran bajas, las cercanas a 0.30 son medias y las correlaciones por arriba de 0.40 son altas.

En ocasiones, es mejor reportar el valor cuadrado de un coeficiente de correlación; en lugar de r, usted verá r^2. Así, si obtenemos r = 0.50, reportamos $r^2 = 0.25$. ¿Por qué se transforma el valor de r? La razón es que la transformación convierte a la r obtenida en porcentaje; el valor representa el porcentaje de va-rianza de una variable que es explicada por la segunda variable. Los valores de r^2 van de 0.00 (0 por ciento) a 1.00 (100 por ciento). En ocasiones, al valor r^2 se le *conoce* como el porcentaje de varianza compartida entre las dos variables. ¿Qué significa esto exactamente? Recuerde el concepto de variabilidad en un conjunto de puntajes: si usted mide el peso de una muestra aleatoria de adultos estadounidenses, observaría la variabilidad en el hecho de que encontraría pesos relativamente bajos y otros relativamente altos. Si usted estudia factores que contribuyen al peso de las personas, querría examinar la relación entre los pesos y los puntajes en la variable de los factores contribuyentes. Una de estas variables podría ser el género. En realidad, la correlación entre el género y el peso es de alrededor de 0.70 (los hombres pesan más que las mujeres). Esto quiere decir que 49 por ciento (elevando al cuadrado 0.70) de la variabilidad de los pesos se explica por la variabilidad del género. Con esto, usted sabe mucho acerca de cómo "explicar" la variabilidad de los pesos, pero aún existe 51 por ciento de varianza sin explicación. La variabilidad puede deberse a otras variables, como el peso de la madre y del padre biológicos, el estrés prenatal, la dieta y el ejercicio. En un mundo ideal, podríamos explicar el 100 por ciento de la variabilidad de los pesos, si tuviéramos suficiente información sobre todas las otras variables que contribuyen al peso de las personas: cada variable añadiría su contribución hasta que toda la variabilidad estuviese explicada.

SIGNIFICACIÓN ESTADÍSTICA

En este capítulo hemos enfatizado la *descripción* de los datos obtenidos en un estudio. Después de describir los datos, usted deseará tomar una decisión con respecto de la significación estadística de los resultados. ¿Es estadísticamente significativa la diferencia entre las medias del grupo con modelo y del grupo sin modelo? La respuesta es, principalmente, cuestión de inferir si los resultados serán los mismos cuando se repita el experimento en varias ocasiones, cada vez con una nueva muestra de participantes en la investigación. La estadística in-ferencial se utiliza para determinar si podemos, de hecho, aseverar que los resultados reflejan lo que sucedería si realizáramos el experimento una y otra vez con múltiples muestras. La significación estadística se aborda en el capítulo 13. El resto de este capítulo trata aspectos adicionales de la correlación.

ECUACIONES DE REGRESIÓN

Las **ecuaciones de regresión** son cálculos utilizados para predecir el puntaje de una persona en una variable, cuando ya se *conoce* su puntaje en otra. Básicamente son "ecuaciones de predicción" que se basan en información conocida acerca de la relación entre las dos variables. Por ejemplo, después de descubrir que el asiento en la clase y la calificación del examen están relacionados, se puede calcular una ecuación de regresión que prediga la calificación del examen de alguno de los individuos, con base únicamente en la información del asiento que ocupa en la clase. La forma general de una ecuación de regresión es

$$Y = a + bX$$

donde *Yes* el puntaje que deseamos predecirles el puntaje conocido, *a* es una constante y *b es* un factor ponderado de ajuste que se multiplica por *X* (es la pendiente de la recta *creada* por medio de esta ecuación). En nuestro ejemplo del asiento y la calificación del examen, calculamos la siguiente ecuación de regresión para los datos:

$$Y = 99 + (-8)X$$

Así, si conocemos el puntaje *X* de la persona (el asiento), al sustituirla en la ecuación podemos predecir cuál será su puntaje Y (la calificación del examen). Si su puntaje *X* es 2 (asiento en la segunda fila), podemos predecir que *Y* = 99 + (- 16), o que su calificación en el examen será de 83. Por medio del uso de ecuaciones de regresión como éstas, las universidades pueden utilizar las calificaciones de la prueba SAT para predecir las calificaciones en la universidad.

Cuando los investigadores se interesan en predecir alguna conducta futura (llamada *variable de criterio),* con base en el puntaje de una persona en alguna

otra variable (llamada *variable predictora),* primero es necesario demostrar que existe una correlación lo suficientemente alta entre la variable de criterio y la predictora. Entonces, la ecuación de regresión proporciona el método para hacer pronósticos únicamente a partir del puntaje de la variable predictora.

CORRELACIÓN MÚLTIPLE

Hasta ahora, nos hemos enfocado en la correlación entre dos variables a la vez. Los investigadores reconocen que una cantidad diferente de variables puede estar relacionada con cierta conducta (se trata del mismo aspecto señalado anteriormente en la explicación de los factores que contribuyen al peso). La técnica llamada **correlación múltiple** se utiliza para combinar diversas variables predictoras e incrementar la precisión del pronóstico de una variable de criterio o de resultado.

La correlación múltiple (simbolizada por *R* para distinguirla de la *r* sencilla) es la correlación entre un conjunto combinado de variables predictoras y una sola variable de criterio. El hecho de tomar en cuenta a todas las variables predictoras, generalmente permite una predicción más precisa que cualquier predictor considerado de manera aislada. Por ejemplo, individuos que solicitan entrar a un curso de posgrado en psicología podrían ser evaluados en un conjunto combinado de variables predictoras, por medio de la correlación múltiple. Las variables predictoras podrían ser: 1) las calificaciones de la licenciatura, 2) la calificación en el examen de admisión general, 3) la calificación en el examen de admisión específico para el curso al que desean ingresar y 4) las cartas de recomendación. Ninguno de estos factores es un predictor perfecto del éxito en los estudios de posgrado, pero esta combinación de variables puede brindar una predicción más precisa. Comúnmente, la correlación múltiple es más alta que la correlación que existe entre cualquiera de las variables predictoras y la variable de criterio o de resultado.

En la práctica real, las predicciones se harían con una extensión de la técnica de la ecuación de regresión estudiada anteriormente. Una ecuación de regresión múltiple puede calcularse de la siguiente manera:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

donde Y es la variable de criterio, X_1 a X_n son las variables predictoras, a es una constante y b_1 a b_n son pesos que se multiplican por los puntajes de las variables predictoras. Por ejemplo, la ecuación de regresión para la admisión al curso de posgrado sería:

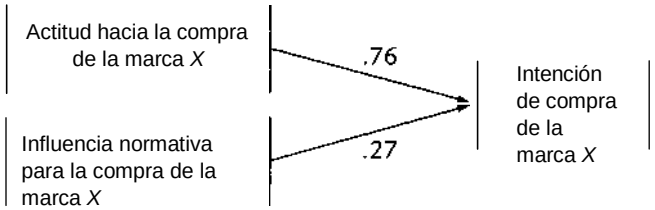
Calificación promedio predicha = a + b_1 (calificaciones de la licenciatura)
+ b_2 (calificación en el examen de admisión general)

$+ b_3$ (calificación en el examen de admisión específico) $+ b_4$ (cartas de recomendación)

Los investigadores utilizan la regresión múltiple para estudiar temas básicos de investigación. Por ejemplo, Ajzen y Fishbein (1980) crearon un modelo llamado "teoría de la acción razonada" que utiliza la correlación y la regresión múltiple para predecir las intenciones específicas del comportamiento (por ejemplo, ir a la iglesia el domingo, comprar cierto producto o entrar a un programa de recuperación del alcoholismo), a partir de dos variables predictoras: 1) la actitud hacia la conducta y 2) la percepción de la presión normativa para realizar la conducta. La actitud es la propia evaluación de la conducta y la presión normativa proviene de otras personas, como los padres y amigos. En un estudio, estos investigadores demostraron que la correlación múltiple entre la intención de comprar una marca de cerveza y los predictores combinados de la actitud y la norma, fue de 0.79. La ecuación de regresión fue la siguiente:

$$\text{Intención} = .76\ (\text{actitud}) + .27\ (\text{norma})$$

Esta ecuación difiere un poco de las que se describieron anteriormente. En la investigación básica no existe un interés en predecir un puntaje exacto (como el resultado de un examen o el promedio de las calificaciones), por lo que en los cálculos matemáticos se puede asumir que todas las variables se miden en la misma escala. Cuando se hace esto, el factor de ponderación refleja la magnitud de la correlación entre la variable de criterio y cada variable predictora. En el ejemplo de la compra de cerveza, el peso del predictor de la actitud es mayor que el peso del predictor de la norma; esto demuestra que, en tal caso, las actitudes son más importantes como predictoras de la intención que las normas. No obstante, en cuanto a otras conductas es probable que las actitudes tengan menor importancia que las normas. También es posible visualizar la ecuación de regresión. En el ejemplo de la compra de cerveza, las relaciones entre variables podrían expresarse como sigue:



Es importante notar que el cuadrado del coeficiente de correlación múltiple (R^2) se interpreta de forma muy similar al cuadrado del coeficiente de correlación (r^2). Esto quiere decir que R^2 indica el porcentaje de variabilidad de la va-

riable de criterio, explicado por el conjunto combinado de variables predictoras. Nuevamente, este valor será mayor que cualquiera de los predictores de manera aislada.

CORRELACIÓN PARCIAL Y EL PROBLEMA DE LA TERCERA VARIABLE

Los investigadores enfrentan el problema de la tercera variable, en estudios no experimentales, cuando alguna tercera variable que no está controlada puede ser responsable de la relación entre las dos variables de interés. Este problema no surge en la investigación experimental, debido a que se controlan todas las variables extrañas, al mantenerlas constantes o por medio de la aleatorización. La técnica denominada **correlación parcial** ofrece una forma de control estadístico de la tercera variable. Una correlación parcial es aquella que existe entre dos variables de interés, donde se ha eliminado total o parcialmente la influencia de la tercera variable.

Suponga que un investigador encuentra que la correlación entre el hacinamiento residencial y el desempeño en una prueba es de -0.50; un mayor hacinamiento está asociado con puntajes más bajos. El investigador sospecha que puede estar operando una tercera variable. La clase social, por ejemplo, podría influir en el grado de hacinamiento y en el desempeño. El uso de correlaciones parciales implica medir a los sujetos en la tercera variable, además de las dos variables principales. Así, el investigador mide a los participantes en todas las variables: el hacinamiento, el desempeño y la clase social.

Cuando se calcula una correlación parcial entre el hacinamiento y el desempeño, separando la clase social, podemos determinar si la correlación original se reduce de manera sustancial. ¿Se disminuye mucho nuestra correlación original de -0.50, cuando se elimina la influencia de la clase social? La figura 12.9 muestra dos correlaciones parciales diferentes; en ambas existe una correlación de -0.50 entre el hacinamiento y el desempeño. Sin embargo, la primera correlación parcial baja hasta -0.09 cuando se controla estadísticamente la clase social y la segunda correlación parcial permanece alta, aun cuando se elimine la influencia de la clase social. El resultado de la correlación parcial depende de la magnitud de las correlaciones entre la tercera variable y las dos variables de mayor interés.

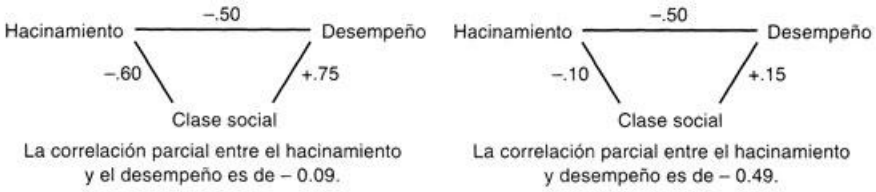


FIGURA 12.9
Dos correlaciones parciales entre el hacinamiento y el desempeño.

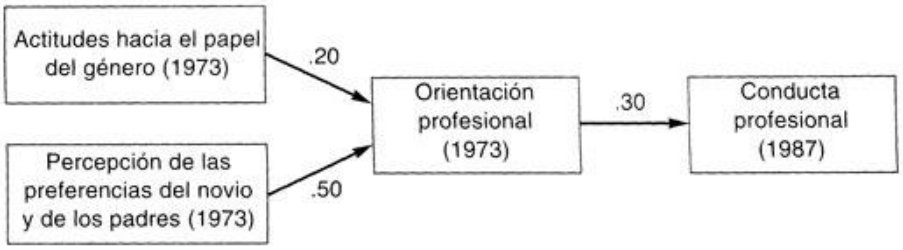


FIGURA 12.10
Modelo estructural basado en los datos de Vincent, Peplau y Hill (1998).

MODELOS ESTRUCTURALES

Avances recientes en la teoría y los métodos estadísticos han producido técnicas para probar **modelos estructurales** de las relaciones entre variables, por medio del método no experimental. A pesar de que estos métodos rebasan los objetivos de este libro, usted debe saber que existen (véase Loehlin, 1998; Maruyama, 1998). Un modelo estructural es un patrón esperado de relaciones entre un con-junto de variables. El patrón propuesto se basa en una teoría sobre la forma en que las variables están relacionadas entre sí, de manera causal. Este método de investigación se *conoce* como *modelamiento estructural o modelamiento de ecuación estructural,* debido a que las técnicas realmente permiten a los investigadores probar qué tanto los datos obtenidos se ajustan a un "modelo estructural" teórico, que describe las relaciones entre variables (asimismo, las operaciones matemáticas que subyacen a la técnica se basan en ecuaciones estructurales).

El diagrama de las actitudes e intenciones, mostrado anteriormente, constituye un modelo estructural muy sencillo. En la figura 12.10 se incluye un modelo un poco más complejo, basado en la misma teoría de la acción razonada. Vincent, Peplau y Hill (1998) estudiaron las aspiraciones profesionales y las actitudes hacia el papel del género de 105 mujeres jóvenes (con una edad promedio de 21 años) en 1973. En esa época, midieron tres variables: 1) las actitudes hacia el papel del género o qué tan positivamente valoraban las mujeres los papeles femeninos "tradicionales" en la sociedad, 2) la percepción de las preferencias de los padres y de los novios por seguir trayectorias ocupacionales "tradicionales" y 3) la orientación ocupacional. Esta última se midió pidiendo a las mujeres que describieran sus planes o intenciones para el futuro: ¿serían principalmente madres y amas de casa, o se percibían involucradas en profesiones de tiempo completo fuera del hogar? Los investigadores pudieron encontrar a las mismas mujeres en 1987, para medir su trayectoria real.

Las flechas que van de una variable a otra describen las rutas obtenidas que relacionan las variables en el modelo; las flechas indican una secuencia causal propuesta, sugerida por la teoría de la acción razonada. Observe que la conducta real se añadió al modelo; las intenciones conductuales influyen de manera directa a la conducta real. Los coeficientes son similares a los pesos derivados

de las ecuaciones de regresión descritas anteriormente. Indican la fuerza de una relación en la escala que ya nos es familiar, de 0.00 a 1.00. Las actitudes hacia el papel del género y la percepción de las preferencias de los padres y novios se relacionan con las inclinaciones ocupacionales. La presión normativa que proviene de los padres y los novios posee un impacto más fuerte que las actitudes. La orientación ocupacional en 1973 está, a su vez, relacionada con la conducta profesional real de 1987. Vincent y sus colaboradores utilizaron técnicas de modelamiento de ecuación estructural para demostrar que los datos obtenidos se ajustaron al modelo presentado en la figura 12.10.

Una herramienta más antigua pero relacionada, para explorar relaciones entre variables, es el **análisis de ruta.** Este se utiliza principalmente para explorar posibles relaciones entre variables, mientras que el modelo de ecuación estructural se emplea para probar el ajuste de los datos a un modelo teórico, o para comparar el ajuste de varios modelos a datos obtenidos. En ambos casos, este tipo de investigación conduce a una mejor comprensión de las redes complejas de relaciones entre variables.

En el siguiente capítulo, pasamos de la descripción de datos a la toma de decisiones sobre la significación estadística. Por supuesto, ambos temas están relacionados; el tema de la magnitud del efecto, descrito en este capítulo, también es muy importante para evaluar la significación estadística.

Términos de estudio

Análisis de ruta
Coeficiente de correlación
Coeficiente de correlación producto-momento de Pearson
Correlación múltiple
Correlación parcial
Desviación estándar
Diagrama de dispersión
Distribución de frecuencias
Ecuaciones de regresión
Escalas de intervalo
Escalas de razón
Escalas nominales
Escalas ordinales
Estadística descriptiva
Gráfica de barras
Gráfica de pastel
Magnitud del efecto
Media
Mediana
Moda
Modelos estructurales
Polígonos de frecuencias
Restricción del rango
Significación estadística
Tendencia central
Variabilidad
Variable de criterio
Variable predictora
Varianza

Preguntas de repaso

1. ¿Cuál es la diferencia entre comparar porcentajes, comparar medias y correlacionar puntajes?
2. ¿Qué es una distribución de frecuencias?

3. Determine la diferencia entre una gráfica de pastel, una de barras y un polígono de frecuencias. Elabore una de cada tipo.
4. ¿Qué es una medida de tendencia central? Establezca las diferencias entre la media, la mediana y la moda.
5. ¿Qué es una medida de variabilidad? Establezca la diferencia entre la desviación estándar y el rango.
6. ¿Qué es un coeficiente de correlación? ¿Qué indican el tamaño y el signo del coeficiente de correlación sobre la relación entre variables?
7. ¿Qué es un diagrama de dispersión?
8. ¿Qué sucede cuando un diagrama de dispersión indica una relación curvilínea?
9. ¿Cuál es la diferencia entre un coeficiente de correlación/magnitud del efecto y el método de correlación?
10. ¿Qué es una ecuación de regresión? ¿De qué forma podría un empleador utilizar una ecuación de regresión?
11. ¿De qué manera incrementa la correlación múltiple la precisión de un pronóstico?
12. ¿Cuál es el propósito de la correlación parcial?
13. Cuando se hace un diagrama de un modelo estructural, ¿qué información brindan las flechas que van de una variable a otra?

Actividades

1. Su periódico o revista favorita constituye una rica fuente de estadística descriptiva, en una diversidad de temas. Examine los periódicos de la semana pasada y cualquier revista en su hogar; describa al menos cinco ejemplos de datos reales presentados. Éstos pueden incluir encuestas, experimentos, datos de negocios e incluso información deportiva.
2. Hill (1990) estudió la correlación entre la calificación del examen final en un curso introductorio de sociología y otras variables como la cantidad de faltas. Se obtuvieron los siguientes coeficientes de correlación r de Pearson con la calificación del examen final:

Calificación promedio general de la licenciatura	.72
Cantidad de faltas	-.51
Horas dedicadas al estudio durante la semana	-.11 (no significativo)
Horas dedicadas al estudio durante los fines de semana	.31

Describa cada correlación y elabore gráficas que muestren la forma general de cada relación. ¿Por qué las horas dedicadas al estudio durante

los fines de semana se correlacionan con las calificaciones, pero no las horas de estudio durante la semana?

3. Pregunte a 20 estudiantes en su universidad la cantidad de materias a las que se inscribieron este semestre, así como el número de horas por semana que trabajan de forma remunerada. Elabore una distribución de frecuencias y calcule la media para cada conjunto de datos. Construya un diagrama de dispersión que muestre la relación entre la carga de clases y las horas de trabajo por semana. ¿Parece existir una relación entre las variables? (Nota: Si existe un problema de restricción del rango en su universidad, debido a que pocos estudiantes trabajan o porque la mayoría de ellos se inscribe aproximadamente en la misma cantidad de materias, plantee preguntas diferentes, como la cantidad de horas a la semana que dedican al estudio y a ver televisión.)
4. Antes del inicio del año escolar, la profesora King revisó la información acumulada de los estudiantes de su clase de cuarto grado. Ella encontró que la desviación estándar de las calificaciones de los estudiantes en la prueba de lectura fue de exactamente 0.00. ¿Qué información le proporciona esto? ¿Cómo se podría utilizar esta información?

13

Comprensión de los resultados de investigación: inferencia estadística

Muestras y poblaciones

Estadística inferencial

Hipótesis nula e hipótesis de investigación

Probabilidad y distribuciones muéstrales
Probabilidad: el caso de la PES
Distribuciones muéstrales Tamaño muestral

Ejemplo: las pruebas *t* y *F*
Prueba *t*
Grados de libertad
Pruebas de una cola y pruebas de dos colas
Prueba *F*
Cálculo de la magnitud del efecto
Intervalos de confianza y significación estadística
Significación estadística: panorama general

Errores tipo I y tipo II
Decisiones correctas
Errores tipo I
Errores tipo II
El contexto cotidiano de los errores tipos I y tipo II

Elección de un nivel de significación

Interpretación de resultados no significativos

Elección de un tamaño de muestra: análisis de la potencia

La importancia de las replicaciones

Significación de un coeficiente *r* de correlación de Pearson

Análisis de datos por computadora

Selección de la prueba de significación adecuada
Una variable independiente con dos grupos
Datos de escala nominal
Datos de escala ordinal
Datos de escalas de intervalo y de razón
Una variable independiente con tres o mas grupos
Datos de escala nominal
Datos de escala ordinal
Datos de escalas de intervalo y de razón
Dos o más variables independientes
Datos de escala nominal
Datos de escala ordinal
Datos de escalas de intervalo y de razón
Términos de estudio
Preguntas de repaso
Actividades

En el capítulo anterior examinamos formas para describir los resultados de un estudio. Además de la estadística descriptiva, los investigadores están interesados en la estadística inferencial. Es necesario inferir si los resultados que se obtuvieron en un estudio en particular, serían iguales si el estudio se repitiera una y otra vez. En este capítulo abordamos los métodos para hacer inferencias.

MUESTRAS Y POBLACIONES

La estadística inferencial es necesaria porque los resultados de un estudio se basan en los datos obtenidos de una sola muestra de participantes de investigación. En raras ocasiones, los investigadores estudian poblaciones enteras; sus hallazgos se fundamentan en datos muéstrales. Además de describir estos datos, deseamos hacer aseveraciones acerca de las poblaciones. ¿Se mantendrán los resultados obtenidos en un estudio en particular si éste se repite varias veces?

En el experimento hipotético descrito en el capítulo 12 (véase la tabla 12.1) se obtuvieron puntuaciones medias de la agresión en las condiciones con modelo y sin modelo. Estas medias son diferentes: los niños que observaron a un modelo agresivo, después se comportaron de forma más agresiva que los niños que no vieron al modelo. La **estadística inferencial** se utiliza para determinar si, de hecho, podemos aseverar que los resultados reflejan lo que sucedería cuando realizáramos el experimento muchas veces con diversas muestras. En esencia, nos preguntamos si podemos inferir que la diferencia entre las medias muéstrales, que se presenta en la tabla 12.1, refleja una verdadera disimilitud entre las medias poblacionales.

En el capítulo 7 abordamos este aspecto, en el tema sobre los datos de encuestas. Una muestra de personas de su estado podría indicarle que el 57 por ciento prefiere al candidato *A* y que el 43 por ciento está en favor del candidato *B.* El reporte dice después que estos resultados tienen una precisión de tres puntos porcentuales, con un nivel de confianza de 95 por ciento. Esto significa que

los investigadores tienen mucha confianza en que, si fueran capaces de estudiar a la población completa y no sólo a una muestra, el porcentaje real de individuos que preferiría al candidato *A* estaría entre 60 por ciento y 54 por ciento, y que el porcentaje que elegiría al candidato *B* sería de entre 46 por ciento y 40 por ciento. En este caso, el investigador podría predecir con gran certeza que el candidato *A* ganará, ya que no existe un traslape en los valores proyectados de la población.

La estadística inferencial nos permite llegar a este tipo de conclusiones sobre la base de datos muéstrales. En el estudio que incluye condiciones con modelo y sin modelo, ¿tenemos confianza en que las medias son lo suficientemente distintas como para inferir que se obtendría la misma diferencia en toda la población?

ESTADÍSTICA INFERENCIAL

Gran parte de la discusión anterior sobre los diseños experimentales se centró en la importancia de asegurarse que los grupos sean equivalentes en todos los aspectos, con excepción de la manipulación de la variable independiente. La equivalencia entre los grupos se logra al controlar de forma experimental, o por medio de la aleatorización, todas las demás variables. Se supone que si los grupos son equivalentes, cualesquiera diferencias en la variable dependiente se deberán al efecto de la variable independiente.

Este supuesto suele ser válido. Sin embargo, también es verdad que la diferencia entre dos grupos casi nunca será de cero. En otras palabras, habrá cierta diferencia entre las medias muéstrales, aun cuando se utilicen todos los principios del diseño experimental. Esto sucede debido a que estamos trabajando con muestras y no con poblaciones. El error aleatorio, o por azar, será responsable de algunas de las diferencias entre las medias, incluso si la variable independiente no tuvo efecto alguno sobre la variable dependiente.

Lo importante es que la diferencia entre las medias muéstrales refleja cualquier verdadera diferencia entre las medias poblacionales (es decir, el efecto de la variable independiente), más cualquier error aleatorio. La estadística inferencial permite que los investigadores hagan inferencias sobre la verdadera diferencia entre las poblaciones, con base en datos muéstrales. Específicamente, la estadística inferencial indica la probabilidad de que la diferencia entre las medias esté reflejando el error aleatorio y no una diferencia real.

HIPÓTESIS NULA E HIPÓTESIS DE INVESTIGACIÓN

La inferencia estadística inicia con una aseveración de la hipótesis nula y una hipótesis de investigación (o alternativa). La **hipótesis nula** afirma, simplemente, que las medias poblacionales son iguales; es decir, que la diferencia observada se debe al error aleatorio. La **hipótesis de investigación** señala que las medias

poblacionales no son iguales. La hipótesis nula establece que la variable independiente no tiene un efecto; la hipótesis de investigación afirma que la variable independiente sí tiene un efecto. En el experimento del modelamiento de la agresión, la hipótesis nula y la de investigación son las siguientes:

H_0 (hipótesis nula): la media poblacional del grupo sin modelo es igual a la media poblacional del grupo con modelo.

H_l (hipótesis de investigación): la media poblacional del grupo sin modelo no es igual a la media poblacional del grupo con modelo.

La lógica de la hipótesis nula es la siguiente: si podemos determinar que la hipótesis nula es incorrecta, entonces aceptamos que la hipótesis de investigación es correcta. La aceptación de la hipótesis de investigación implica que la variable independiente tuvo un efecto sobre la variable dependiente.

La hipótesis nula se utiliza debido a que constituye una aseveración muy precisa; las medias poblacionales son exactamente iguales. Esto nos permite conocer con exactitud la probabilidad de la ocurrencia de los resultados, si la hipótesis nula es correcta. Esta precisión no es posible con la hipótesis de investigación, de manera que inferimos que la hipótesis de investigación es correcta únicamente si rechazamos la hipótesis nula. La hipótesis nula se rechaza cuando existen muy pocas probabilidades de que los resultados obtenidos se deban al error aleatorio. A esto se refiere la **significación estadística:** un resultado significativo es aquel que tiene muy poca probabilidad de ocurrir si las medias poblacionales son iguales. En términos más sencillos, la significación indica que existen pocas probabilidades de que la diferencia entre las medias muéstrales obtenidas se deba al error aleatorio. La significación es, entonces, cuestión de probabilidad.

PROBABILIDAD Y DISTRIBUCIONES MUÉSTRALES

La **probabilidad** es la posibilidad de que ocurra algún suceso o resultado. Todos empleamos con frecuencia las probabilidades en la vida cotidiana. Por ejemplo, si usted dice que hay muchas probabilidades de que obtenga diez de calificación en este curso, quiere decir que es posible que logre este resultado. Su afirmación de probabilidad está basada en información específica, como sus calificaciones en los exámenes. El pronóstico del clima dice que hay 10 por ciento de probabilidades de que llueva hoy; esto significa que la posibilidad de lluvia es muy baja. Un apostador calcula la probabilidad de que un caballo en particular gane una carrera, basado en los registros pasados de dicho caballo.

La probabilidad se utiliza para hacer inferencias estadísticas de forma muy parecida. Buscamos especificar la probabilidad de que un acontecimiento (en este caso, una diferencia entre las medias de la muestra) ocurra, si no existen diferencias en la población. La pregunta es: ¿cuál es la probabilidad de obtener este resultado si únicamente opera el error aleatorio? Si la probabilidad es muy

baja, rechazamos la posibilidad de que el error aleatorio sea el único responsable de la diferencia obtenida entre las medias.

Probabilidad: el caso de la PES

La aplicación de la probabilidad en las inferencias estadísticas puede comprenderse de manera intuitiva por medio de un sencillo ejemplo. Suponga que un amigo afirma tener la capacidad de PES (percepción extrasensorial). Usted decide probar a su amigo con un conjunto de cinco cartas que se utilizan comúnmente en la investigación de PES; cada carta contiene un símbolo diferente. En la prueba de PES usted observa cada carta, piensa en el símbolo y su amigo le dice el símbolo que usted está pensando. En su experimento real, usted incluye diez ensayos; presenta cada una de las *cinco* cartas dos veces en un orden aleatorio. Su tarea es saber si las respuestas de su amigo reflejan error aleatorio (adivinación) o si indican que está ocurriendo algo más que un error aleatorio. La hipótesis nula de su estudio es que únicamente ocurre error aleatorio. La hipótesis de investigación es que el número de respuestas correctas demuestra algo más que el azar o adivinación (observe, sin embargo, que aceptar la hipótesis de investigación podría significar que su amigo posee capacidades de PES, aunque también podría significar que las cartas estaban marcadas, que usted le daba pistas a su amigo al pensar en los símbolos, entre otras posibilidades).

Usted puede determinar fácilmente el número de respuestas correctas esperadas, si la hipótesis nula es correcta. Por adivinación, una de cada cinco respuestas (20 por ciento) debe ser correcta. En diez ensayos, según la hipótesis nula, se esperan dos respuestas correctas. Si en el experimento real se obtienen más de dos respuestas correctas (o menos), ¿podría concluir que los datos obtenidos reflejan error aleatorio o algo más que la mera adivinación?

Suponga que su amigo obtiene tres respuestas correctas. Entonces usted probablemente concluiría que está adivinando, debido a que reconocería que hay una alta probabilidad de obtener tres respuestas correctas, *aun cuando sólo se esperan dos respuestas correctas bajo la hipótesis nula.* Usted espera a la larga que, si realizara este experimento con el sujeto muchas veces, exactamente dos respuestas en diez ensayos sean correctas. No obstante, es muy probable que en una muestra de diez ensayos los resultados se desvíen un poco de las dos respuestas correctas esperadas.

Sin embargo, suponga que su amigo logra siete respuestas correctas. Podría concluir que los resultados indican algo más que error aleatorio en esta muestra de diez observaciones. Esta conclusión estaría basada en su juicio intuitivo de que es muy poco probable un resultado con el 70 por ciento de respuestas correctas, cuando únicamente se espera 20 por ciento. En este punto, usted decidiría rechazar la hipótesis nula y afirmar que el resultado es significativo. Un resultado significativo es aquel que tiene muy pocas probabilidades de ocurrir, si la hipótesis nula es correcta.

¿Qué tan poco probable debe ser un resultado para decidir que es significativo? La decisión se determina antes de reunir los datos. La probabilidad re-

querida para la significación se denomina *nivel alfa.* El nivel alfa de probabilidad que más se utiliza es 0.05. Se considera que los resultados de un estudio son significativos cuando hay una probabilidad igual o menor que 0.05 de obtener los resultados; es decir, que existen únicamente 5 posibilidades en 100 de que los resultados se deban al error aleatorio en una muestra de la población. Si hay muy pocas probabilidades de que el error aleatorio sea el responsable de los resultados obtenidos, se rechaza la hipótesis nula.

Distribuciones muéstrales

Quizás usted haya juzgado de forma intuitiva que es muy poco probable obtener siete respuestas correctas en los diez ensayos. Por fortuna, no tenemos que confiar en nuestras impresiones para determinar las probabilidades de distintos resultados. La tabla 13.1 muestra la probabilidad de obtener, en realidad, cada uno de los resultados posibles en el experimento de la PES en diez ensayos, con una hipótesis nula que establece una expectativa de 20 por ciento de respuestas correctas. Un resultado con dos respuestas correctas tiene la probabilidad más alta de ocurrencia. Además, como sugiere la intuición, un resultado con tres respuestas correctas es muy probable, pero uno con siete es altamente improbable. Las probabilidades que se presentan en la tabla 13.1 se derivaron de una distribución de probabilidad llamada *distribución binomial;* todas las decisiones con respecto a la significación estadística se basan en distribuciones de probabilidad de este tipo, las cuales se denominan *distribuciones muéstrales.* La distribución muestral se fundamenta en el supuesto de que la hipótesis nula es verdadera; en el ejemplo de la PES, la hipótesis nula establece que la persona está adivinando y que, por lo tanto, obtendrá 20 por ciento de respuestas correctas.

TABLA 13.1 Probabilidad exacta de cada resultado posible del experimento de la PES con 10 ensayos.

Numero de respuestas correctas	Probabilidad
10	.00000+
9	.00000+
8	.00007
7	.00079
6	.00551
5	.02642
4	.08808
3	.20133
2	.30199
1	.26844
0	.10737

Una distribución como éstas supone que si usted realizara el estudio con el mismo número de observaciones una y otra vez, el hallazgo más frecuente sería de 20 por ciento. No obstante, debido al posible error aleatorio en cada muestra, existe cierta probabilidad asociada con otros resultados. Los resultados cercanos al valor esperado de 20 por ciento, según la hipótesis nula, son muy probables. Sin embargo, resultados distintos a lo esperado son cada vez menos probables, si la hipótesis nula es correcta. Cuando los resultados que obtiene son altamente improbables, y usted seleccionó la muestra de la distribución especificada por la hipótesis nula, concluye que la hipótesis nula es incorrecta. En lugar de concluir que sus resultados muéstrales reflejan una desviación aleatoria del esperado 20 por ciento, usted decide que la hipótesis nula es incorrecta. Es decir, concluye que no obtuvo la muestra de la distribución muestral especificada en la hipótesis nula. En el caso del ejemplo de la PES, usted decide que sus datos provienen de una distribución muestral distinta en la que, si probara a la persona de manera repetida, la mayoría de los resultados se acercarían al producto obtenido de siete respuestas correctas.

Todas las pruebas estadísticas se fundamentan en distribuciones muéstrales para determinar la probabilidad de que los resultados sean consistentes con la hipótesis nula. Cuando es altamente improbable que la hipótesis nula sea correcta (generalmente una probabilidad de 0.05 o menos), el investigador decide rechazar la hipótesis nula y, por lo tanto, aceptar la hipótesis de investigación.

Tamaño muestral

El ejemplo de la PES también hace énfasis en la importancia del tamaño de la muestra —el número total de observaciones— para determinar la significación estadística. Suponga que usted prueba a su amigo en 100 ensayos, en lugar de 10, y observa 30 respuestas correctas. Como esperaba dos respuestas correctas en 10 ensayos, ahora esperaría que 20 de 100 fueran correctas. No obstante, el resultado de 30 de 100 respuestas correctas tiene menos posibilidades de ocurrir que un resultado de tres de cada 10. Esto se debe a que, con más observaciones, usted tiene mayores posibilidades de obtener un estimativo preciso del valor poblacional verdadero. Así, conforme se incrementa el tamaño de la muestra, existe mayor confianza en que el resultado será realmente diferente de lo esperado de acuerdo con la hipótesis nula.

EJEMPLO: LAS PRUEBAS *t* Y *F*

Las distintas pruebas estadísticas nos permiten emplear la probabilidad para decidir si rechazamos la hipótesis nula. En esta sección estudiaremos las pruebas *t y F*. La prueba *t* se utiliza principalmente para examinar si dos grupos difieren entre sí de manera significativa. En el experimento hipotético de los efectos de un modelo sobre la agresión, la prueba *t* es apropiada porque deseamos saber si la media del grupo sin modelo difiere de la media del grupo con

modelo. La prueba *F* es una prueba estadística más general que puede emplearse para saber si existen diferencias entre tres o más grupos, o para evaluar los resultados de diseños factoriales (estudiados en el capítulo 10).

Para utilizar una prueba estadística, primero debe especificar las hipótesis nula y de investigación que está evaluando. Anteriormente describimos las hipótesis nula y de investigación del experimento del modelamiento. También se debe especificar el nivel de significación que se empleará para decidir el rechazo de la hipótesis nula; es decir, el nivel alfa. Como se señaló, los investigadores generalmente utilizan un nivel de significación de 0.05.

Prueba t

En la figura 13.1 se presenta la distribución muestral de todos los valores posibles de *t* (esta distribución, en particular, es del tamaño muestral utilizado en el experimento hipotético del modelamiento y la agresión; el tamaño muestral fue de 20, con diez participantes en cada grupo). Esta distribución muestral tiene una media de 0 y una desviación estándar de 1. Refleja todos los resultados esperados posibles si comparamos las medias de los dos grupos y si la hipótesis nula es correcta.

Para emplear esta distribución con la finalidad de evaluar nuestros datos, necesitamos calcular un valor de *t* a partir de los datos obtenidos y evaluar la *t* resultante en términos de la distribución muestral de *t*, que está basada en la hipótesis nula. Si la *t* obtenida tiene una baja probabilidad de ocurrencia (0.05 o menos), entonces se rechaza la hipótesis nula.

El valor *t* es un cociente de dos aspectos de los datos, la diferencia entre las medias grupales y la variabilidad dentro de los grupos. El cociente puede describirse de la siguiente forma:

$$t = \frac{\text{diferencia grupal}}{\text{variabilidad dentro de los grupos}}$$

La diferencia grupal está, sencillamente, entre las medias obtenidas; según la hipótesis nula, se espera que esta diferencia sea 0. El valor de *t* se incrementa conforme aumenta la diferencia entre las medias muéstrales obtenidas. Observe que la distribución muestral de *t* asume que no existen diferencias entre las medias poblacionales; por lo tanto, el valor esperado de í, de acuerdo con la hipótesis nula, es 0. La variabilidad dentro de los grupos es la cantidad de variabilidad de las puntuaciones con respecto a la media. El denominador de la fórmula de *t* es esencialmente un indicador de la cantidad de error aleatorio en la muestra. En el capítulo 12 establecimos que s_y la desviación estándar, y s^2, la varianza, son indicadores de la cantidad de desviación de las puntuaciones a partir de la media grupal.

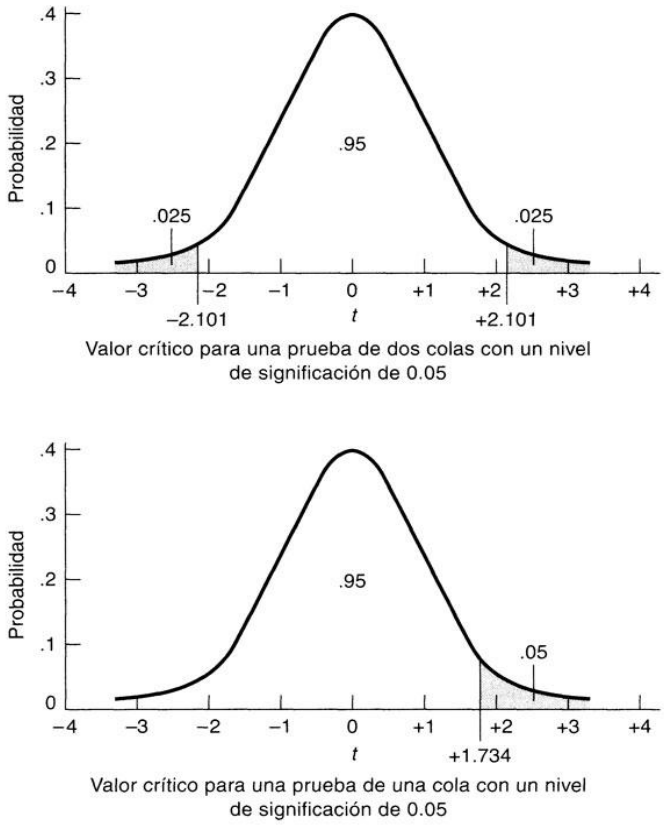


FIGURA **13.1**
Distribución muestral de los valores de t con 18 grados de libertad.

Un ejemplo concreto del cálculo de una prueba *t* ayudará aclarar estos conceptos. La fórmula de la prueba *t* para dos grupos con el mismo número de participantes en cada grupo es:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

El numerador de la fórmula es la diferencia entre las medias de los dos grupos. En el denominador, primero se divide la varianza *(s^2)* de cada grupo entre el número de sujetos en el grupo y después se suman. Luego se calcula la raíz cuadrada del resultado; esto convierte la cifra de una raíz cuadrada (la varianza). en una desviación estándar. Finalmente, se calcula el valor de *t* dividiendo la diferencia de las medias entre esta desviación estándar. Si aplicamos la fórmula a los datos de la tabla 12.1, obtenemos:

$$t = \frac{5.2 - 3.1}{\sqrt{\frac{1.289}{10} + \frac{1.433}{10}}}$$

$$= \frac{2.1}{\sqrt{.1289 + .1433}}$$

$$= 4.025$$

Así, el valor de *t* calculado con los datos es de 4.025. ¿Se trata de un resultado significativo? Si se analizan los resultados con un programa de cómputo, nos indicarían de inmediato la probabilidad de obtener un valor *t* de esta magnitud, con un tamaño muestral total de 20. Sin embargo, si usted no tiene acceso a un programa de este tipo, puede consultar la tabla de "valores críticos" de *t,* como la que se muestra en la tabla C.3, del apéndice C. En el apéndice B explicaremos con detalle el uso de las tablas adjuntas. Antes de continuar, usted debe saber que el resultado obtenido es significativo. Con un nivel de significación de 0.05, el valor crítico de la distribución muestral de *t* es 2.101. Cualquier valor de *t* mayor o igual que 2.101 tiene una probabilidad de 0.05 o menos de ocurrir, bajo los supuestos de la hipótesis nula. Puesto que el valor obtenido es mayor que el valor crítico, podemos rechazar la hipótesis nula y concluir que la diferencia entre las medias en la muestra refleja una verdadera diferencia en la población.

Grados de libertad

Quizás se pregunte cómo se eligió el valor crítico en la tabla. Para utilizar la tabla, primero debe determinar los **grados de libertad** de la prueba (el término grados de libertad se abrevia *gl).* Cuando se comparan dos medias, los grados de libertad son iguales a $N_1 + N_2 — 2$, o el número total de participantes en los grupos menos la cantidad de grupos. En nuestro experimento, los grados de libertad serían 10 + 10 — 2 = 18. Los grados de libertad implican la cantidad de puntuaciones que pueden variar libremente, una vez que se conocen las medias. Por ejemplo, si la media de un grupo es 6.0 y hay un total de *cinco* puntuaciones en el grupo, entonces existen cuatro grados de libertad; cuando tenga cualesquiera cuatro puntuaciones, la quinta se *conoce* debido a que la media debe continuar siendo 6.0.

Pruebas de una cola y pruebas de dos colas

En la tabla, usted debe elegir una *t* crítica para la situación en la que su hipótesis de investigación 1) especifica una dirección de la diferencia entre los grupos (por ejemplo, el grupo 1 será mayor que el grupo 2) o 2) no especifica una dirección o diferencia (por ejemplo, el grupo 1 diferirá el grupo 2). En ambas situaciones se utilizan valores críticos con ciertas diferencias. La primera situación es una prueba de una cola y la segunda es una prueba de dos colas.

Se puede visualizar este aspecto cuando se observa la distribución muestral de los valores de *t* para 18 grados de libertad, como se presenta en la figura 13.1. Se observa que, el valor 0.00 es el que se espera con mayor frecuencia. Valores mayores o menores que cero tienen menores posibilidades de ocurrir. La primera distribución presenta la lógica de una prueba de dos colas. Utilizamos la cifra de 2.101 como valor crítico de *t,* con un nivel de significación de 0.05, en la que no se estableció la dirección de la diferencia. Este valor crítico es el punto más allá del cual cae el 2.5 por ciento de los valores positivos y el 2.5 por ciento de los valores negativos de *t* (por lo tanto, una probabilidad total de 0.05, combinada de las dos "colas" de la distribución muestral). La segunda distribución ejemplifica una prueba de una cola. Si se hubiese predicho una diferencia direccional, el valor crítico habría sido 1.734. Este es el valor a partir del cual cae el 5 por ciento de los valores en una sola "cola" de la distribución. La especificación de una prueba de una o dos colas depende de si el estudio se diseñó originalmente para probar una hipótesis direccional.

Prueba F

El **análisis de varianza** o **prueba** Fes una extensión de la prueba *t.* El análisis de varianza es un procedimiento estadístico más *general* que la prueba *t.* Cuando un estudio tiene una sola variable independiente con *dos* grupos, *F y t* son vir-tualmente idénticas, el valor de *F* es igual a t^2 en esta situación. Sin embargo, el análisis de varianza también se utiliza cuando la variable independiente tiene más de dos niveles y cuando se empleó un diseño factorial con dos o más variables independientes. Así, la prueba *F* es apropiada para el diseño experimental más simple, pero también para los diseños más complejos estudiados en el capítulo 10. Incluimos la prueba *t* porque la fórmula nos permite demostrar con facilidad la relación de la diferencia grupal y la variabilidad dentro de los grupos con los resultados de la prueba estadística. Pero en la práctica el análisis de varianza es el procedimiento más utilizado. En el apéndice B se presentan los cálculos necesarios para realizar una prueba *F.*

La estadística *F* es un cociente de dos tipos de varianza: la varianza sistemática y la varianza del error (de ahí el término *análisis de varianza).* La **varianza sistemática** es la desviación de las medias grupales a partir de la gran media o puntuaciones medias de todos los individuos en todos los grupos. La varianza sistemática es pequeña cuando la diferencia entre las medias grupales es pequeña, y se incrementa conforme aumentan las diferencias entre las medias grupa-

les. La **varianza del error** es la desviación de las puntuaciones individuales en cada grupo, a partir de sus medias grupales respectivas. Algunos de los términos de investigación que se utilizan en lugar de varianza sistemática y del error son *varianzd entre grupos* y *varianza dentro de los grupos.* La varianza sistemática es la variabilidad de las puntuaciones entre los grupos, y la varianza del error es la variabilidad de las puntuaciones dentro de los grupos. Mientras más grande es el cociente i[7], existen mayores posibilidades de que los resultados sean significativos.

Cálculo de la magnitud del efecto

En el capítulo 12 estudiamos el concepto de la magnitud del efecto. Además de saber si hay un efecto estadísticamente significativo de la variable independiente, es útil conocer la magnitud del efecto. Por lo tanto, debemos calcular un estimativo de la magnitud del efecto. Para una prueba t, los cálculos son

$$r \text{ de la magnitud del efecto} = \sqrt{\frac{t^2}{t^2 + gl}}$$

Donde *gl* son los grados de libertad. Así, utilizando el valor obtenido de *t*, 4.025, con 18 grados de libertad, obtenemos

$$r \text{ de la magnitud del efecto} = \sqrt{\frac{(4.025)^2}{(4.025)^2 + 18}} = \sqrt{\frac{16.201}{34.201}} = .688$$

Este valor es un tipo de coeficiente de correlación que va de 0.00 a 1.00. Para obtener información adicional sobre los cálculos de la magnitud del efecto, revise Rosentahl (1991). Aquí se aplica la misma distinción entre ryr^2 que se hizo en el capítulo 12.

Otro estimativo de la magnitud del efecto, que se utiliza cuando se comparan dos medias, es la *d* de Cohén. La *d* de Cohén expresa el tamaño del efecto en términos de unidades de desviación estándar. Una *d* con valor de 1.0 indica que las medias están a una distancia de una desviación estándar; una *d* de 0.5 indica que las medias están separadas por 0.5 desviaciones estándar. El valor de *d* es mayor que el valor correspondiente de r. Ambas estadísticas proporcionan información sobre el tamaño de la relación entre las variables que se estudian. Podrá notar que ambos estimativos de la magnitud del efecto tienen un valor de 0.00 cuando no existe relación. La r posee un valor máximo de 1.00, pero *d* no tiene un valor máximo.

Intervalos de confianza y significación estadística

En el capítulo 7 describimos los intervalos de confianza. Después de obtener un valor muestral, se puede calcular un intervalo de confianza. Un intervalo de va-

lores define el rango más probable de valores poblacionales reales. El intervalo se asocia en un nivel de confianza: un intervalo de confianza de 95 por ciento indica que tenemos una seguridad de 95 por ciento de que el valor poblacional cae dentro del rango; un intervalo de 99 por ciento proporciona mayor certeza, aunque el rango de valores sería más grande.

Se puede obtener un intervalo de confianza para cada una de las medias del experimento de la agresión. Los intervalos de confianza de 95 por ciento de las dos condiciones son:

	Valor muestral obtenido	*Valor poblacional bajo*	*Valor poblacional alto*
Grupo con modelo	5.20	4.39	6.01
Grupo sin modelo	3.10	2.24	3.96

Es muy importante examinar los intervalos de confianza para lograr una mayor comprensión del significado de los datos obtenidos. A pesar de que las medias muéstrales proporcionan el mejor estimativo de los valores poblacionales, usted podrá ver el rango probable de los posibles valores. El tamaño del intervalo se relaciona con el tamaño de la muestra y con el nivel de confianza. Conforme el tamaño muestral se incrementa, el intervalo de confianza se reduce. Esto se debe a que las medias muéstrales obtenidas con muestras de mayor tamaño tienen más posibilidades de reflejar la media poblacional. En segundo lugar, un alto nivel de confianza se asocia con un intervalo mayor. Si desea asegurarse de que el intervalo contenga la media poblacional verdadera (por ejemplo, un intervalo de confianza de 99 por ciento), necesita incluir más posibilidades. Observe que los intervalos de confianza de 95 por ciento de las dos medias no se traslapan. Esto debe servirle como indicador de que la diferencia es estadísticamente significativa. De hecho, el examen de los intervalos de confianza es una forma alternativa de considerar la significación estadística. La hipótesis nula establece que la diferencia entre las medias poblacionales es de 0.00. Sin embargo, si restara todas las medias en el intervalo de confianza de 95 por ciento de la condición sin modelo, de todas las medias de la condición con el modelo, ninguna de estas diferencias incluiría el valor 0.00. Podemos tener mucha confianza en que la hipótesis nula debe rechazarse.

Significación estadística: panorama general

La lógica que subyace al uso de las pruebas estadísticas está en la teoría estadística. No obstante, existen algunos conceptos generales que deben ayudarlo a comprender lo que hace cuando realiza una prueba estadística. Primero, la meta de la prueba es permitirle tomar decisiones respecto a la confiabilidad de los resultados obtenidos; Usted desea tener confianza en que obtendrán resultados similares si realiza el estudio una y otra vez. En segundo lugar, el nivel de signi-

ficación (nivel alfa) que emplea indica qué tan confiado desea estar al tomar la decisión. Un nivel de significación de 0.05 indica que usted tiene una seguridad de 95 por ciento de la confiabilidad de sus hallazgos; aunque hay una posibilidad de 5 por ciento de equivocarse. ¡Existen pocas certezas en la vida! En tercer lugar, tiene mayores posibilidades de obtener resultados significativos cuando tiene una muestra grande, ya que las muestras grandes proporcionan mejores estimaciones de los valores poblacionales verdaderos. Finalmente, tiene mayores posibilidades de obtener resultados significativos cuando la magnitud del efecto es grande; es decir, cuando las diferencias entre los grupos son grandes y la variabilidad de los puntajes dentro de los grupos es pequeña.

En lo que resta del capítulo profundizaremos en estos temas. Examinaremos las implicaciones de la toma de decisiones respecto a la significación de los resultados, la forma para determinar el nivel de significación y la manera de interpretar resultados que no son significativos. Después, daremos algunos lineamientos para seleccionar la prueba estadística adecuada para diversos diseños de investigación.

ERRORES TIPO I Y TIPO II

La decisión de rechazar la hipótesis nula se basa en probabilidades y no en certezas. Es decir, la decisión se toma sin *conocer* directamente la verdadera situación de la población. Por lo tanto, la decisión podría ser incorrecta; pueden resultar errores del uso de la estadística inferencial.

En la figura 13.2 se muestra una matriz de decisión. Observe que existen dos decisiones posibles: 1) rechazar la hipótesis nula o 2) aceptar la hipótesis nula. También existen dos verdades posibles acerca de la población: 1) la hipótesis nula es verdadera o 2) la hipótesis nula es falsa. En resumen, como se muestra en la matriz de decisión, existen dos tipos de decisiones correctas y dos tipos de errores.

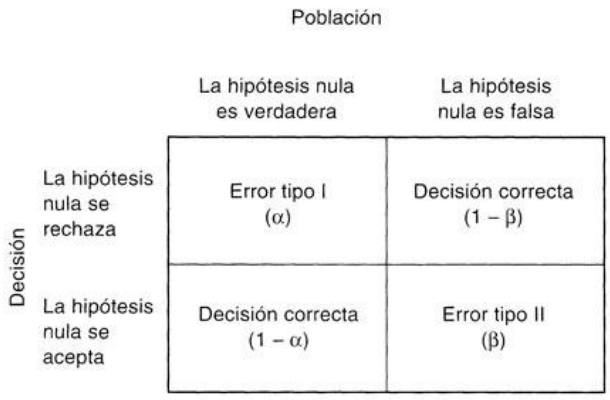


FIGURA 13.2
Matriz de decisiones de los errores tipo I y tipo **I**

Decisiones correctas

Tomamos una decisión adecuada cuando rechazamos la hipótesis nula y la hipótesis de investigación es verdadera en la población. Aquí se decide que las medias poblacionales no son iguales y, de hecho, esto es verdad en la población. Ésta es la decisión que deseamos tomar al iniciar el estudio.

La otra decisión correcta implica aceptar la hipótesis nula cuando ésta es verdadera en la población: las medias poblacionales son iguales.

Errores tipo I

Cometemos un **error tipo I** cuando rechazamos la hipótesis nula y ésta, en realidad, es verdadera. La decisión implica que las medias poblacionales no son iguales cuando, de hecho, sí lo son. Los errores tipo I ocurren cuando, sólo por azar, obtenemos un valor grande de *t* o *F*. Por ejemplo, aun cuando es sumamente improbable obtener un valor *t* de 4.025 si las medias poblacionales son iguales (menos de 5 posibilidades en 100), esto *puede* suceder. Cuando obtenemos un valor *t* tan grande como éste por el azar, *de manera incorrecta* decidimos que la variable independiente tuvo un efecto.

Las probabilidades de cometer un error tipo I están determinadas por la elección del nivel de significación o alfa (alfa puede aparecer como la letra griega a). Cuando el nivel de significación para decidir el rechazo de la hipótesis nula es de 0.05, la probabilidad de cometer un error tipo I (alfa) es de 0.05. Si se rechaza la hipótesis nula, existen 5 posibilidades en 100 de que la decisión sea incorrecta. La probabilidad de cometer un error tipo I puede modificarse disminuyendo o aumentando el nivel de significación. Por ejemplo, si utilizamos un nivel alfa menor, de 0.01, existen menores posibilidades de cometer este tipo de error. Con un nivel de significación de 0.01, la hipótesis nula se rechaza únicamente cuando la probabilidad de tener el resultado es de 0.01 o menor, si la hipótesis nula es correcta.

Errores tipo II

Cometemos un **error tipo II** cuando se acepta la hipótesis nula, aun cuando la hipótesis de trabajo es verdadera en la población. Las medias poblacionales no son iguales, pero los resultados del experimento no condujeron a la decisión de rechazar la hipótesis nula.

Las investigaciones deben diseñarse de tal forma que la probabilidad de un error tipo II (a esta probabilidad se le llama beta o (3) sea relativamente baja. La probabilidad de cometer un error tipo II está relacionada con tres factores. El primero es el nivel de significación (alfa). Si establecemos un nivel de significación demasiado bajo para disminuir las posibilidades de un error tipo I, se incrementan las posibilidades de cometer un error tipo II. En otras palabras, si hacemos que sea muy difícil el rechazo de la hipótesis nula, se incrementan las posibilidades de que la aceptemos erróneamente. El segundo factor es el tama-

ño muestral. Es más probable detectar diferencias verdaderas cuando el tamaño de la muestra es grande. El tercer factor es la magnitud del efecto; si la magnitud del efecto es grande, el error tipo II es improbable. Sin embargo, una pequeña magnitud del efecto podría no ser significativa con una muestra pequeña.

El contexto cotidiano de los errores tipo I y tipo II

La matriz de decisión que se utiliza en el análisis estadístico podría aplicarse al tipo de decisiones que las personas deben tomar con frecuencia en la vida cotidiana. Por ejemplo, considere las decisiones que toma un jurado en un juicio criminal. Como sucede con la estadística, se debe tomar una decisión con base en las evidencias: ¿es inocente o culpable el acusado? No obstante, la decisión reside en los jurados y no necesariamente refleja la verdad: que la persona realmente es inocente o culpable.

En la figura 13.3 se muestra la matriz de decisión del jurado. Para continuar el símil con la decisión estadística, suponga que la hipótesis nula establece que el acusado es inocente (es decir, la máxima de que una persona es inocente hasta que se pruebe lo contrario). Así, rechazar la hipótesis nula significa decidir que el acusado es culpable, y aceptarla implica decidir que es inocente. La matriz de decisión también muestra que la hipótesis nula puede ser realmente verdadera o falsa. Existen dos tipos de decisiones correctas y dos tipos de errores, como los que se describieron en las decisiones estadísticas. Se comete un error tipo I al declarar que el acusado es culpable, cuando en realidad es inocente; se comete un error tipo II si se le declara inocente, cuando en realidad es culpable. En nuestra sociedad, los errores tipo I generalmente se consideran más graves que los errores tipo II. Por lo tanto, antes de declarar culpable al individuo, se le pide al jurado que se asegure que es culpable "más allá de cualquier duda razonable" o que considere que "es mejor que cien personas culpables estén libres que declarar culpable a una persona inocente".

La decisión de un médico de operar o n o a un paciente es otro ejemplo del funcionamiento de la matriz de decisión. Esta matriz se presenta en la figura 13.4. Aquí, la hipótesis nula establece que no se requiere de una operación. La decisión implica rechazar la hipótesis nula *y* llevar a cabo la operación o aceptar la hipótesis nula y omitir la cirugía. En realidad, el cirujano se enfrenta a dos posibilidades: o la cirugía es innecesaria (la hipótesis nula es verdadera) o el paciente morirá sin la operación (un caso dramático en el que la hipótesis nula es falsa). ¿Cuál error es más grave en este caso? La mayoría de los médicos considerarían que es más grave no operar a un paciente que realmente necesita la cirugía —cometer un error tipo II— que el error tipo I, el de realizar la cirugía a un individuo que en realidad no la necesita.

Un último ejemplo de lo usos de una matriz de decisión implica la importante decisión de casarse. Si la hipótesis nula establece que la persona es "incorrecta" para usted, y la situación verdadera es que es "incorrecta" o "correcta", debe decidir si sigue adelante y se casa con ella. Puede tratar de elaborar una ma-

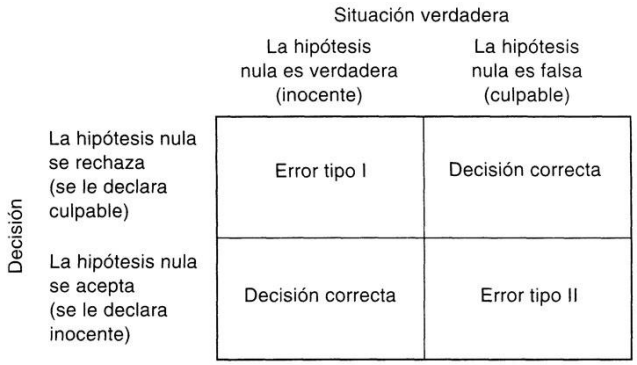


FIGURA 13.3
Matriz de decisión de un jurado.

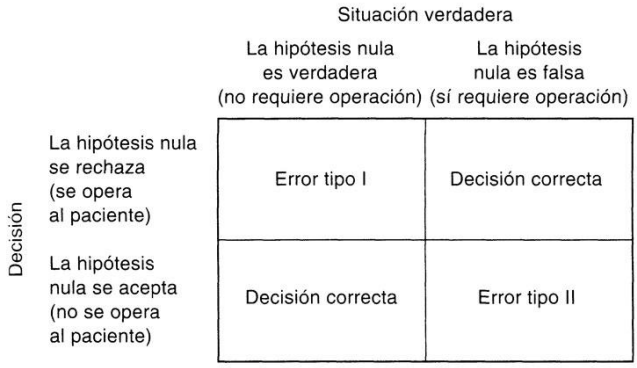


FIGURA 13.4
Matriz de decisión de un médico.

triz de decisión para este problema en particular. ¿Cuál de los errores es más costoso: un error tipo I o un error tipo II?

ELECCIÓN DE UN NIVEL DE SIGNIFICACIÓN

Por tradición, los investigadores utilizan un nivel de significación de 0.05 o de 0.01, para tomar la decisión de rechazar la hipótesis nula. Si existe una proba-bilidad menor que 0.05 o 0.01 de que los resultados se deban al error aleatorio, se dice que los resultados son significativos. Sin embargo, los niveles de sig-nificación de 0.05 o de 0.01 no son mágicos. El nivel de significación que se elija únicamente especifica la probabilidad de cometer un error tipo I si se rechaza la hipótesis nula. El nivel de significación elegido generalmente depende de las consecuencias de cometer un error tipo I, con relación a un error tipo II. Como

se señaló anteriormente, es más grave que un jurado cometa un error tipo I que un error tipo II; sin embargo, para un médico puede ser más grave cometer un error tipo II.

Los investigadores suelen considerar que las consecuencias de cometer un error tipo I son más graves que las asociadas con un error tipo II. Si se rechaza la hipótesis nula, el investigador puede publicar los resultados en una revista científica y otros autores podrían publicar sus resultados en libros de texto o artículos de periódicos o revistas. Los investigadores no desean confundir a las personas o arruinar su reputación al publicar resultados que no son confiables y que no pueden replicarse. Por lo tanto, buscan protegerse en contra de la posibilidad de cometer un error tipo I, por medio de un nivel de significación muy bajo (0.05 o 0.01). A diferencia de las consecuencias de la publicación de resultados falsos, las consecuencias de un error tipo I no se consideran tan graves.

Así, los investigadores son muy cuidadosos para evitar errores tipo I cuando publican sus resultados. Sin embargo, en ciertas circunstancias, un error tipo I no suele ser tan grave. Por ejemplo, si usted realiza un estudio piloto o exploratorio, sus resultados se emplearían principalmente para decidir si sus ideas de investigación valen la pena. En esta situación, sería erróneo ignorar datos potencialmente importantes al utilizar un nivel de significación muy conservador. En la investigación exploratoria, un nivel de significación de 0.25 podría ser el más apropiado para decidir si continúa con la investigación. Recuerde que el nivel de significación elegido y las consecuencias de los errores tipos I y II están determinados por el uso que se dará a los resultados.

INTERPRETACIÓN DE RESULTADOS NO SIGNIFICATIVOS

A pesar de que los términos "aceptar la hipótesis nula" son convenientes, es importante reconocer que generalmente los investigadores no se interesan en la aceptación de la hipótesis nula. Las investigaciones se diseñan para demostrar que existe una relación entre variables, y no para demostrar que las variables no están relacionadas.

Es muy importante tomar en cuenta que la decisión de aceptar la hipótesis nula, cuando un estudio no muestra resultados significativos, representa un problema, debido a que los resultados negativos o no significativos son difíciles de interpretar. Es probable que los resultados de un solo estudio no sean significativos, aun cuando realmente exista una relación entre las variables en la población; éste es un error tipo II. En ocasiones, la causa de un error tipo II reside en los procedimientos empleados en el experimento. Por ejemplo, un investigador podría obtener resultados no significativos cuando da instrucciones incomprensibles a los participantes, al manejar una manipulación muy débil de la variable independiente o cuando utiliza una medida dependiente poco confiable y poco sensible. En lugar de concluir que las variables no están relacionadas, quizás un estudio más cuidadoso descubriría la relación entre las variables.

También debemos considerar las causas estadísticas de un error tipo II. Re-cuerde que las probabilidades de cometer un error tipo II se ven afectadas por el nivel de significación (alfa), el tamaño muestral y la magnitud del efecto. Así, es más probable obtener resultados no significativos si se es demasiado cauto en la elección del nivel alfa. Si el investigador utiliza un nivel de significación de 0.001 para decidir el rechazo de la hipótesis nula, no hay muchas posibilidades de co-meter un error tipo I; no obstante, sí es posible cometer un error tipo II, puesto que el investigador disminuyó las posibilidades de rechazar erróneamente la hipótesis nula. En otras palabras, existen más probabilidades de dejar pasar por alto un resultado significativo cuando el nivel de significación es muy bajo.

Un error tipo II también puede resultar de un tamaño muestral que es demasiado pequeño como para detectar una relación real entre variables. Un principio general establece que a mayor tamaño muestral, mayores son las posibilidades de obtener un resultado significativo. Esto se debe a que las muestras grandes proporcionan estimativos más precisos de la población real que las muestras pequeñas. En cualquier estudio, el tamaño muestral puede ser demasiado pequeño como para permitir la detección de un resultado significativo.

La tercera razón de un hallazgo no significativo reside en una magnitud pequeña del efecto. Es muy difícil detectar efectos muy pequeños si no se tiene una muestra grande. En general, el tamaño muestral debe ser lo suficientemente grande como para encontrar un efecto real, aun cuando éste sea pequeño.

El hecho de que un efecto muy pequeño pueda ser estadísticamente significativo plantea otra situación. Una muestra muy grande puede permitir que los investigadores encuentren una diferencia significativa entre medias; aun cuando esta diferencia sea estadísticamente significativa, podría tener muy poca significación *practica*. Por ejemplo, si un nuevo tratamiento psiquiátrico costoso reduce de manera significativa la estancia hospitalaria de un promedio de 60 a 59 días, quizás no sea práctico utilizar el tratamiento, a pesar de la evidencia de su eficacia; el día adicional de hospitalización es menos costoso que el tratamiento. Sin embargo, existen otras circunstancias en las que un tratamiento con un efecto de pequeña magnitud tiene una gran significación práctica. Esto suele ocurrir cuando una población muy grande se ve afectada por un tratamiento poco costoso. Suponga que una sencilla política de flexibilidad de tiempo con los empleados reduce la rotación de personal a 1 por ciento anual; esta situación no parece tener un gran efecto. Pero si una empresa normalmente cambia a dos mil empleados cada año, y el costo de entrenamiento de cada uno es de $10 000, la empresa ahorra $200 000 anuales con el nuevo procedimiento. Esta cantidad puede tener una significación práctica para la empresa.

Lo más importante de esto es que no debemos aceptar la hipótesis nula sólo debido a que los resultados no sean significativos. Los resultados no significativos no necesariamente indican que la hipótesis nula es correcta. No obstante, debe haber circunstancias en las que podemos aceptar la hipótesis nula y concluir que dos variables no están relacionadas. Frick (1995) describe varios criterios que pueden emplearse para tomar la decisión de aceptar la hipótesis

nula. Por ejemplo, debemos buscar estudios diseñados adecuadamente, con medidas dependientes sensibles y evidencias de la verificación de la manipulación que indiquen que la manipulación de la variable independiente tuvo el efecto deseado. Además, la investigación debe incluir una muestra lo suficientemente grande como para eliminar los problemas de una muestra demasiado pequeña. A su vez, la evidencia de que las variables no están relacionadas debe provenir de diversos estudios. Bajo estas circunstancias, existe la justificación para concluir que no hay una relación.

ELECCIÓN DE UN TAMAÑO DE MUESTRA: ANÁLISIS DE LA POTENCIA

En el capítulo 9 señalamos que los investigadores con frecuencia seleccionan el tamaño muestral basados en lo que se acostumbra hacer en cierta área de investigación. Un método alternativo implica seleccionar el tamaño muestral con base en las probabilidades deseadas para rechazar de manera correcta la hipótesis nula. A esta probabilidad se le *conoce* como la **potencia** de la prueba estadística y, obviamente, está relacionada con la probabilidad de un error tipo II:

$$\text{Potencia} = 1 - p \text{ (error tipo II)}$$

Anteriormente, indicamos que la probabilidad de cometer un error tipo II está relacionada con el nivel de significación (alfa), con el tamaño muestral y con la magnitud del efecto. Estadísticos como Cohén (1988) han creado procedimientos para determinar el tamaño muestral con base en estos factores. La tabla 13.2 indica el tamaño muestral total necesario para un experimento con dos grupos y un nivel de significación de 0.05. En la tabla, la magnitud del efecto va de 0.10 a 0.50, y se incluyen las potencias deseadas de 0.80 y 0.90. Para que magnitudes pequeñas del efecto sean significativas, se requiere de muestras grandes. Las potencias altas demandan un tamaño muestral mayor; esto se debe a que deseamos "garantizar" con mayor certeza que los resultados serán estadísticamente significativos. Los investigadores suelen utilizar una potencia de entre 0.70 y 0.90 cuando emplean este método de determinación del tamaño muestral. Existen varios programas de cómputo que facilitan a los investigadores los cálculos necesarios para determinar los tamaños muéstrales con base en estimativos de la magnitud del efecto, el nivel de significación y la potencia deseada.

Es probable que usted nunca necesite realizar un análisis de la potencia. Sin embargo, debe reconocer la importancia de este concepto. Si un investigador está estudiando una relación con una correlación de la magnitud del efecto de 0.20, se requiere de una muestra bastante grande para lograr una significación estadística en el nivel de 0.05. En esta situación, un tamaño muestral demasiado pequeño puede producir un hallazgo no significativo.

TABLA 13.2 Tamaño muestral total necesario para detectar una diferencia significativa en una prueba t.

r de la magnitud del efecto	Potencia = 0.80	Potencia = 0.90
.10	786	1 052
.20	200	*266*
.30	88	116
.40	52	68
.50	28	36

Nota: las magnitudes del efecto son correlaciones; basado en pruebas de dos colas.

LA IMPORTANCIA DE LAS REPLICACIONES

A lo largo de esta explicación acerca de los análisis estadísticos, nos hemos enfocado en los resultados de una sola investigación. ¿Cuáles son las medias y la desviación estándar? ¿Es la diferencia entre las medias estadísticamente significativa? Si los resultados son significativos, usted puede concluir que probablemente se obtengan una y otra vez si el estudio se repite. Ahora tenemos un marco de referencia para comprender los resultados del estudio. Sin embargo, tenga en mente que los científicos no dan demasiada importancia a los resultados de un solo estudio. La mayor comprensión de cualquier fenómeno proviene de los resultados de numerosos estudios que investigan las mismas variables. En lugar de inferir los valores poblacionales a partir de una investigación, podemos revisar los resultados de varios estudios que replican investigaciones previas (véase Cohén, 1994). La importancia de la replicación es un concepto central del capítulo 14.

SIGNIFICACIÓN DE UN COEFICIENTE r DE CORRELACIÓN DE PEARSON

En el capítulo 12 estudiamos que el coeficiente de correlación *r* de Pearson se emplea para describir la fuerza de la relación entre dos variables, cuando ambas tienen propiedades de una escala de intervalo o de razón. No obstante, es necesario ver si la correlación es estadísticamente significativa. En este caso, la hipótesis nula establece que la correlación poblacional verdadera es de 0.00; es decir, las dos variables no están relacionadas. ¿Qué sucede si se obtiene una correlación de 0.27 (más o menos)? Una prueba de significación estadística permite tomar la decisión de rechazar la hipótesis nula y concluir que la correlación poblacional verdadera es, de hecho, mayor que 0.00. Para hacer esto, se realiza una prueba *t* que compara el coeficiente obtenido con la correlación de

0.00 establecida en la hipótesis nula. Los procedimientos para el cálculo de la r de Pearson y para determinar la significación se incluyen en el apéndice B.

ANÁLISIS DE DATOS POR COMPUTADORA

Aun cuando es posible calcular estadísticas con una calculadora, por medio de las fórmulas incluidas en este capítulo, en el capítulo 12 y en el apéndice B, la mayoría de los análisis de datos se realiza por medio de programas de cómputo. Los paquetes sofisticados de análisis estadísticos facilitan el cálculo de estadísticas de cualquier conjunto de datos; las estadísticas descriptivas e inferenciales se obtienen con rapidez, los cálculos son precisos y en los resultados aparece información sobre la significación estadística. Las computadoras también facilitan la presentación gráfica de los datos.

Algunos de los programas estadísticos de cómputo más conocidos son SPSS, SAS, Minitab, Systat y BMDP; también podrá utilizar otros programas en su universidad. Numerosas personas realizan la mayoría de sus análisis estadísticos por medio de un programa de hojas de cálculo como Microsoft Excel. Usted necesitará aprender los detalles específicos de los sistemas de cómputo utilizados en su universidad. Ningún programa es mejor que otro, todos difieren en la presentación de los resultados, en los procedimientos específicos necesarios para alimentar los datos y en las instrucciones de programa para realizar las pruebas. No obstante, los procedimientos generales para hacer análisis son bastante similares en todos los programas de estadística.

El primer paso para el análisis es la introducción de los datos. Suponga que desea introducir los datos de la tabla 12.1, del experimento del modelamiento y la agresión. Los datos se introducen en columnas. Es más fácil si consideramos los datos para el análisis por computadora como una matriz con renglones y columnas. Los renglones de la matriz representan los datos de cada participante de investigación; las columnas contienen las puntuaciones de cada participante en una o más medidas, y tal vez se necesite una columna adicional para indicar un código que identifique la condición en que participó cada individuo (por ejemplo, grupo 1 o grupo 2). En la figura 13.5 se muestra una matriz de datos en el programa SPSS para Windows. Los números en la columna "grupos" indican si el individuo participó en el grupo 1 (con modelo) o en el 2 (sin modelo), y las cifras en la columna "condagre" son los actos agresivos indicados en la tabla 12.1.

Es probable que otros programas requieran métodos un poco diferentes para ingresar los datos. En Excel, por ejemplo, generalmente es más fácil establecer una columna separada para cada grupo, tal como se muestra en la figura 13.5.

El siguiente paso incluye las instrucciones para el análisis estadístico. Nuevamente, cada programa utiliza pasos distintos para realizar el análisis; la mayoría ofrece diversas opciones en un menú. Una vez que el análisis se ha completado, aparecen los resultados del procedimiento estadístico realizado. Es necesario que aprenda a interpretar los resultados. La figura 13.5 muestra la hoja de resultados de una prueba *t* realizada en Excel.

	grupo	condagre
1		3
2		4
3		5
4		5
5		5
6	2	1
7	2	2
8	2	2
9	2	3
10	2	3
11	1	5
12	1	6
13	2	3
14	2	4
15	2	4

Matriz de datos del SPSS para Windows

	A	B
1	Con modelo	Sin modelo
2	3	1
3	4	2
4	5	2
5	5	3
6	5	3
7	5	3
8	6	4
9	6	4
10	6	4
11	7	5
12		
13		

Método de Excel para la introducción de datos

Prueba t: dos muestras con supuestas varianzas iguales

	Con modelo	*Sin modelo*
Media	5.200	3.100
Varianza	1.289	1.433
observaciones	10.000	10.000
Varianza agrupada	1. 361	
diferencia entre medias hipotetizadas	0.000	
gl	18.000	
Estado t	4.025	
P (T<=t) una cola	0.000	
t critica una cola	1. 734	
P ('T<=t) dos colas	0.001	
t critica dos colas	2.101	

Resultados de una prueba t por medio de Excel

FIGURA 13.5

Ejemplo de la captura y los resultados de los datos de la figura 12.1 (experimento de mdelamiento).

Cuando empiece a aprender el uso de un programa de análisis estadístico, una buena idea es practicar con algunos datos de un libro de estadística para asegurarse de obtener los mismos resultados. Esto garantizará que sepa introducir los datos y solicitar el análisis estadístico de forma adecuada.

SELECCIÓN DE LA PRUEBA DE SIGNIFICACIÓN ADECUADA

Se han elaborado muchas pruebas estadísticas para distintos diseños de investigación. La prueba de significación adecuada depende del diseño utilizado y del tipo de escala de medición empleada para estudiar las variables. En el capítulo 12 estudiamos que los valores de las variables tienen propiedades de escalas de medición. Existen cuatro tipos de escalas: 1) nominal, en la que los valores no poseen propiedades numéricas; 2) ordinal, en la que los valores únicamente tienen un significado de orden; 3) de intervalo, en la que los valores se ordenan en un continuo con intervalos iguales, y 4) de razón, en la que existe un cero verdadero en la escala de valores. Esta sección no ofrece lineamientos para seleccionar una prueba en todas las situaciones, sino únicamente las pruebas adecuadas para algunos de los diseños más comunes.

Una variable independiente con dos grupos

Datos de escala nominal Cuando los individuos en dos grupos se miden por medio de una escala nominal, lo más adecuado es una *prueba chi cuadrada*. Esta prueba se utiliza con todos los datos nominales. Consulte un libro de estadística para aprender la aplicación de la fórmula.

Datos de escala ordinal Si se utiliza un diseño de grupos independientes, la prueba de significación estadística más adecuada es la *U* de Mann-Whitney. Para un diseño de medidas repetidas o participantes apareados, se utiliza la *T* de Wilco-xon o prueba de los signos. Los procedimientos de estas pruebas pueden encontrarse en muchos libros de estadística (véase Siegel y Castellan, 1988).

Datos de escalas de intervalo y de razón Para diseños de grupos independientes se emplea la prueba *t* descrita en este capítulo o el análisis de varianza de un factor. Los cálculos del análisis de varianza se describen en el apéndice B. Para diseños de medidas repetidas o grupos apareados debe emplearse una prueba *t* con un procedimiento ligeramente diferente o un análisis de varianza de medidas repetidas. Los cálculos para el análisis de varianza de medidas repetidas se incluyen en el apéndice B.

Una variable independiente con tres o más grupos

Datos de escala nominal En esta situación se utiliza la prueba chi cuadrada.

Datos de escala ordinal Para un diseño de grupos independientes la prueba *H de Kruskal-Wallace* es la más adecuada. Para diseños de medidas repetidas se utiliza la prueba *T de Friedman*. Consulte un libro de estadística para estudiar los detalles del cálculo de estas pruebas.

Datos de escalas de intervalo y de razón Aquí se utiliza un análisis de varianza de un factor. Los procedimientos para el cálculo de esta prueba, para diseños de grupos independientes y de medidas repetidas, se incluyen en el apéndice B.

Dos o más variables independientes

Datos de escala nominal Nuevamente, la prueba chi cuadrada es adecuada para datos de escala nominal. En el apéndice B encontrará la fórmula para la chi cuadrada en la que los participantes están clasificados en dos variables, ambas en escala nominal.

Datos de escala ordinal No existe una prueba estadística adecuada.

Datos de escalas de intervalo y de razón Para diseños factoriales con dos variables independientes se utiliza un análisis de varianza de dos factores. El análisis de varianza puede aplicarse a diseños con cualquier cantidad de variables independientes y es adecuado para diseños de grupos independientes, de medidas repetidas o mixtos, tanto con grupos independientes y con variables de medidas repetidas.

Ahora ya sabe cómo generar ideas de investigación, realizar investigaciones para probar sus ideas y evaluar la significación estadística de sus resultados. En el capítulo final estudiaremos aspectos de la generalización de los hallazgos de investigación, más allá de las circunstancias específicas en las que se realizó el estudio.

Términos de estudio

Análisis de varianza (prueba *F)*
Distribución muestral
Error tipo I
Error tipo II
Estadística inferencial
Grados de libertad
Hipótesis de investigación
Hipótesis nula
Nivel alfa
Potencia
Probabilidad
Prueba chi cuadrada
Prueba *t*
Significación estadística
Varianza del error
Varianza sistemática

Preguntas de repaso

1. Establezca la diferencia entre la hipótesis nula y la hipótesis de investigación. ¿Cuándo decide el investigador rechazar la hipótesis nula?

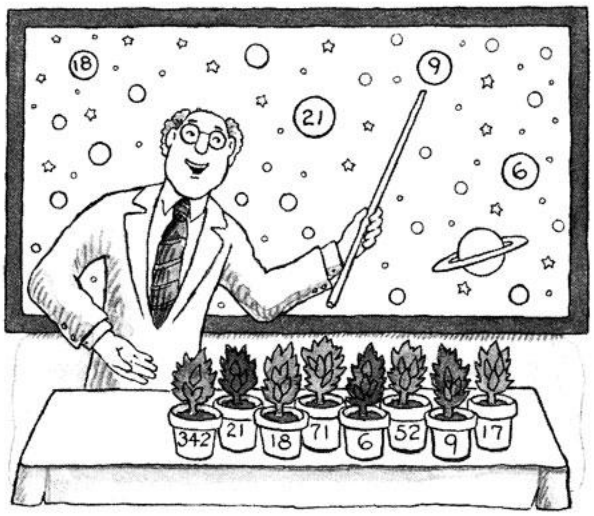
2. ¿Qué es la significación estadística?
3. ¿Cuáles factores son más importantes para determinar si los resultados obtenidos serán significativos?
4. ¿Cuál es la diferencia entre un error tipo I y un error tipo II? ¿Por qué el nivel de significación es la probabilidad de cometer un error tipo I?
5. ¿Qué factores están involucrados en la elección de un nivel de significación?
6. ¿Qué aspectos influyen en la probabilidad de un error tipo II?
7. ¿Cuál es la diferencia entre significación estadística y significación práctica?
8. Comente las razones por las que un investigador podría obtener resultados no significativos.

Actividades

1. En un experimento, a un grupo de participantes de investigación se le dan diez páginas de material para determinar los errores en una prueba de lectura. Otro grupo realiza la misma prueba de lectura en la pantalla de una computadora. La variable dependiente es la cantidad de errores detectados en un periodo de cinco minutos. Se utiliza un nivel de significación (alfa) de 0.05 para evaluar los resultados.
   a) ¿Qué prueba estadística utilizaría?
   b) ¿Cuál es la hipótesis nula? ¿Y la hipótesis de investigación?
   c) ¿Cuál es el error tipo I? ¿Y cuál el error tipo II?
   d) ¿Cuál es la probabilidad de cometer un error tipo I?
2. En el estudio del profesor Dre, el promedio del número de errores detectados en las condiciones de impresión y de computadoras fue de 38.4 y 13.2, respectivamente; esta diferencia no fue estadísticamente significativa. Cuando el profesor Seuss realizó el mismo experimento, las medias de los dos grupos fueron 21.1 y 14.7, pero la diferencia fue estadísticamente significativa. Explique cómo pudo suceder esto.
3. Suponga que trabaja para el departamento de servicio social infantil de su municipio. Su trabajo es investigar casos de posible negligencia o abuso infantil. Después de reunir evidencias, que pueden provenir de diversas fuentes, usted debe decidir si permite que el niño permanezca en su hogar o ponerlo bajo custodia de protección. Especifique la hipótesis nula y la de investigación en esta situación. ¿Cuál sería un error tipo I y un error tipo II? ¿Cuál error es más grave en esta situación, el tipo I o el tipo II? ¿Por qué?
4. Un investigador estudió las actitudes hacia los individuos en sillas de ruedas. La pregunta fue: ¿reaccionará la gente de manera diferente ante

una persona a quien perciben confinada en forma temporal a la silla de ruedas que ante una persona que tiene una discapacidad permanente? Los participantes fueron asignados aleatoriamente a dos grupos. Cada uno de los individuos de un grupo trabajó en diversas tareas con un cómplice en silla de ruedas; los miembros del otro grupo trabajaron con el mismo cómplice en una silla de ruedas, pero que en esta ocasión tenía una pierna enyesada. Al finalizar la sesión, los participantes completaron un cuestionario relativo a sus reacciones en el estudio. Una de las preguntas fue: "¿estaría dispuesto a trabajar con su compañero de prueba en el futuro para un trabajo de clase?"; las únicas opciones de respuesta fueron "sí" y "no". ¿Cuál sería la prueba de significación adecuada para este experimento? ¿Podría dar una crítica de la variable dependiente? Si cambiara la variable dependiente, ¿afectaría su elección de la prueba de significación? Si su respuesta es sí, ¿cómo?

14



Generalización de los resultados

Generalización a otras poblaciones
Estudiantes universitarios
Voluntarios
Aspectos de género
Lugar
La generalización como interacción estadística
En defensa de los estudiantes universitarios y de las ratas

Consideraciones culturales

Generalización a otros experimentadores

Pretest y generalización

Generalización a partir de ambientes de laboratorio
Realismo mundano y experimental
Beneficios recíprocos de la investigación de laboratorio y de campo

La importancia de las replicaciones
Replicaciones directas
Replicaciones conceptuales

Evaluación de la generalización mediante revisiones de la literatura y meta-análisis

Uso de la investigación para mejorar las condiciones de vida

Términos de estudio

Preguntas de repaso

Actividades

En este capítulo abordaremos el tema de la generalización de los resultados de investigación. Un estudio se realiza con una muestra y procedimientos particulares. ¿Pueden generalizarse los resultados a otras poblaciones, o a otras formas de manipulación o medición de las variables? Recuerde que la validez interna se refiere a la posibilidad de inferir que existe una relación causal entre variables. La validez externa es la amplitud en que los hallazgos pueden ser generalizados.

GENERALIZACIÓN A OTRAS POBLACIONES

Aun cuando el investigador asigna de manera aleatoria a los participantes a las condiciones experimentales, pocas veces los selecciona al azar de la población general. En los capítulos 7 y 9, señalamos que los individuos que participan en investigaciones psicológicas se seleccionan debido a que están disponibles, y la población más accesible es la de estudiantes universitarios, específicamente de primero y segundo años, que estudian cursos introductorios de psicología para cumplir un requisito educativo general. También pueden formar parte de una universidad en particular, ser voluntarios o pueden ser principalmente hombres o principalmente mujeres. ¿Se limitan nuestros hallazgos de investigación a este tipo de sujetos, o podemos generalizar los resultados a una población más amplia? Después de abordar estos temas, examinaremos el aspecto más extenso de la cultura.

Estudiantes universitarios

Smart (1966) encontró que más del 70 por ciento de los artículos publicados entre 1962 y 1964 en *el Journal of Experimental Psycbology* y en el *Journal ofAbnor-mal and Social Psychology* incluyeron estudiantes universitarios como participantes. Sears (1986) reportó porcentajes similares en 1980 y 1985, en diversas revistas de psicología social. El problema potencial consiste en que dichos estudios utilizan una población sumamente restringida. Sears señala que la mayoría de los estudiantes son de primero y segundo años, y pertenecen a clases de cursos introductorios de psicología. Por lo tanto, suelen ser muy jóvenes y poseen las características de la adolescencia tardía: una identidad aún en desarrollo, actitudes sociales y políticas fluctuantes, una gran necesidad de aprobación de los compañeros y relaciones inestables con los mismos. Asimismo, son inteligentes, tienen grandes habilidades cognoscitivas y saben ganar la aprobación de la autoridad (habiendo tenido éxito en un ambiente escolar para entrar a la universidad). Así, lo que conocemos sobre "la gente en general" puede, en realidad, estar limitado a un grupo sumamente selecto y poco común.

El problema de la poca representatividad de los sujetos no se limita a la in-vestigación con seres humanos. Una gran cantidad de investigación con animales se basa únicamente en ratas blancas. ¿Por qué? En parte, como señala Beach (1950), porque "las ratas son resistentes, poco costosas, fáciles de criar y porque

se adaptan bien a la vida del laboratorio". De esa manera, igual que los estudiantes de primero y segundo años, son fáciles de reclutar en un *campus* universitario.

Voluntarios

Generalmente, los investigadores deben pedir a las personas que participen voluntariamente en un estudio. En muchas universidades, se requiere que los estudiantes de cursos introductorios de psicología participen voluntariamente en experimentos o que realicen un proyecto alternativo. Si usted estudia poblaciones distintas a los estudiantes universitarios, dependerá aún más de los voluntarios; por ejemplo, al pedir a las personas en una reunión de una asociación de propietarios de casas que participen en un estudio sobre la interacción marital, o al realizar una investigación en Internet en la que los individuos deben abrir su página y después aceptar participar en el estudio. Las investigaciones indican que los voluntarios difieren en muchos aspectos de los individuos que no son voluntarios (Rosenthal y Rosnow, 1975). Por ejemplo, los voluntarios suelen tener un mayor nivel educativo, mayor necesidad de aceptación y ser más sociables; además, tienden a pertenecer a un nivel socioeconómico más alto.

Asimismo, distintos tipos de personas se ofrecen como voluntarios para distintos tipos de experimentos. En las universidades, es posible que haya un tablero con los nombres de diversos estudios. Es probable que las personas que deciden participar en un estudio titulado "solución de problemas", sean distintas a las que se interesan por uno llamado "interacción en grupos pequeños". Las evidencias disponibles indican que el nombre del estudio influye en los tipos de individuos que deciden participar (Hood y Back, 1971; Silverman y Margulis, 1973).

Aspectos de género

En ocasiones, los investigadores estudian bien sea hombres o bien mujeres (o una tasa muy desproporcionada de hombres y mujeres), simplemente debido a que es conveniente o a que los procedimientos parecen ser más adecuados para hombres o para mujeres. Sin embargo, dadas las posibles diferencias entre hombres y mujeres, los resultados de este tipo de estudios no pueden ser generali-zables (Denmark, Russo, Frieze y Sechzer, 1988). Denmark y sus colaboradores ofrecen un ejemplo de estudios sobre prácticas de anticoncepción que utilizan mujeres únicamente, debido al estereotipo de que sólo las mujeres son responsables de dicha práctica. También señalan varias formas del sesgo por género en la investigación psicológica, incluyendo la confusión entre género y edad o nivel de empleo y la selección de medidas de respuestas con estereotipos de género. La solución es estar conscientes de posibles diferencias por género e incluir tanto a hombres como a mujeres en las investigaciones. Además, es importante reconocer las formas en que hombres y mujeres pueden interpretar de manera distinta

las manipulaciones de la variable independiente o las preguntas planteadas en un cuestionario.

Lugar

Los participantes de un lugar pueden ser diferentes de los participantes de otras partes. Por ejemplo, los estudiantes de la UCLA pueden diferir de los estudiantes de una universidad estatal cercana, quienes, a su vez, pueden ser distintos de los estudiantes de una universidad comunitaria. Es probable que las personas de Iowa sean diferentes a las de la ciudad de Nueva York. Por lo tanto, los hallazgos obtenidos con los estudiantes de un tipo de ambiente educativo o de una región geográfica quizás no puedan generalizarse a la gente de otros ambientes o regiones.

La generalización como interacción estadística

El problema de la generalización puede verse como una interacción en un diseño factorial (véase el capítulo 10). Una interacción ocurre cuando la relación entre variables existe en una condición pero no en otra, o cuando la naturaleza de la relación es distinta en una condición que en otra. Así, si usted cuestiona la generalización de un estudio que incluye únicamente hombres, usted está sugiriendo que existe una interacción entre el género y la variable independiente. Suponga, por ejemplo, que un estudio examina la relación entre el hacinamiento y la agresión entre hombres, y reporta que el hacinamiento está asociado con altos niveles de agresión. Usted podría cuestionarse si los resultados pueden generalizarse a las mujeres.

En la figura 14.1 se muestran cuatro posibles resultados de un estudio hipotético sobre el hacinamiento y la agresión, que incluyó hombres y mujeres. En cada gráfica, se mantiene la misma relación entre el hacinamiento y la agresión en los hombres. En la gráfica A, no hay una interacción; es decir, el comportamiento de hombres y mujeres es virtualmente idéntico. De este modo, los resultados del estudio original con hombres, podrían generalizarse a las mujeres. En la gráfica B, tampoco hay una interacción; el efecto del hacinamiento es idéntico en los hombres y las mujeres. Sin embargo, en esta gráfica los hombres son más agresivos que las mujeres. A pesar de que esta diferencia es interesante, no es un problema para la generalización, dado que la relación general entre el hacinamiento y la agresión está presente tanto para hombres como para mujeres.

Las gráficas C y D muestran interacciones. En ambas, los resultados originales de los hombres no pueden generalizarse a las mujeres. En la gráfica C, no existe una relación entre el hacinamiento y la agresión en las mujeres. En la gráfica D, la interacción nos indica la existencia de una relación positiva entre el hacinamiento y la agresión en los hombres, pero una relación negativa en las mujeres. La gráfica D describe los resultados de varios estudios (por ejemplo, Freedman, Levy, Buchanan y Pnce, 1972).

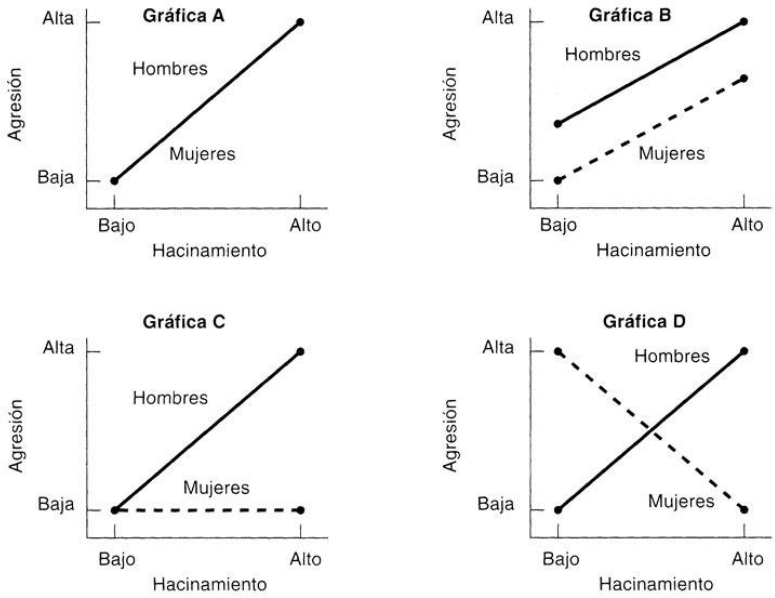


FIGURA 14.1
Resultados de un experimento hipotético sobre hacinamiento y agresión.
Nota: la presencia de una interacción indica que los resultados de los hombres no pueden generalizarse a las mujeres.

Al incluir el tipo de sujeto como variable del estudio, los investigadores pueden resolver problemas de la generalización que surgen del uso de distintas poblaciones. Cuando se incluyen variables como el género, la edad o el grupo étnico en el diseño del estudio, los resultados pueden analizarse para determinar si existen efectos de interacción, como los que se ilustran en la figura 14.1.

En defensa de los estudiantes universitarios y de las ratas

Es fácil criticar las investigaciones con base en las características de los sujetos, aunque la crítica por sí misma no implica que el estudio sea necesariamente un fracaso. Aun cuando debemos preocuparnos por los problemas potenciales de hacer generalizaciones a partir de poblaciones únicas, como los estudiantes universitarios (véase Sears, 1986), también debemos tener dos cosas en mente cuando pensamos en este tema. Primero, las críticas respecto al uso de un tipo particular de sujetos en un estudio, como los estudiantes universitarios o cualquier otro grupo, deben respaldarse con buenas razones que justifiquen que no

es posible encontrar la relación con otro tipo de sujetos. Después de todo, los estudiantes universitarios *son* seres humanos, y los investigadores no deben ser culpados por no preocuparse por generalizar con base en un tipo particular de sujetos, si no hay buenas razones para tal preocupación. Además, los estudiantes universitarios son diversos y representan a la sociedad como un todo. En segundo lugar, recuerde que la replicación de estudios de investigación proporciona una salvaguarda en contra de la posibilidad de la generalización limitada. Los estudios se replican en otras universidades utilizando distintas mezclas de estudiantes, y muchos hallazgos que se establecen primero en estudiantes universitarios, posteriormente se aplican a otras poblaciones como niños, ancianos e individuos de otros países. Asimismo, a pesar de que las ratas son resistentes y poco costosas, el valor de los estudios que emplean ratas ha quedado demostrado por las investigaciones que aplican estos hallazgos a los problemas humanos. Por ejemplo, las investigaciones sobre el reforzamiento de ratas y pichones se han aplicado para modificar el comportamiento, comprender la personalidad y estudiar la conducta de elección en seres humanos.

CONSIDERACIONES CULTURALES

Hasta hace pocos años, los participantes de la mayoría de los experimentos realizados con estudiantes universitarios en Estados Unidos eran individuos de raza blanca, porque esto reflejaba la población predominante en las universidades. Sin embargo, en la actualidad la mayoría de las muestras de estudiantes universitarios son de orígenes étnicos diversos. Con esto se increméntala validez externa y es mucho más fácil comparar grupos étnicos para examinar diferencias y similitudes grupales. A finales de los ochenta, menos del 10 por ciento de los estudios de psicología social, incluían comparaciones de dos o más grupos culturales (Pepitone y Triandis, 1987). Actualmente, existe un mayor interés por el estudio de distintas culturas (Miller, 1999).

Gran parte de la investigación cultural se centra en la identificación de similitudes y diferencias que pueden existir en la personalidad y en otras características psicológicas, así como en la forma en que los individuos de diferentes culturas responden a los mismos ambientes (Matsumoto, 1994). Por ejemplo, Cunningham, Roberts, Barbee, Druen y Wu (1995) realizaron una serie de estudios para examinar la percepción del atractivo físico de las mujeres en distintas culturas. En una investigación, estudiantes estadounidenses blancos y estudiantes asiáticos e hispanos recién llegados a Estados Unidos, calificaron el atractivo de los rostros de mujeres asiáticas, negras, hispanas y blancas, presentados en fotografías. Las calificaciones fueron muy similares entre los grupos; la correlación promedio fue de 0.93. Sin embargo, se descubrieron ciertas diferencias culturales. Por ejemplo, en un estudio en el que hombres estadounidenses negros y blancos calificaron el rostro y el cuerpo, los grupos mostraron calificaciones similares para los rostros, pero difirieron en su apreciación del cuerpo. Además, mientras que la mayoría consideró más atractivas a las mujeres con

una apariencia "sexualmente madura", los asiáticos no se vieron influidos por la madurez sexual.

Este tipo de investigaciones nos brinda información acerca de la generalidad de los hallazgos entre los grupos culturales. Miller (1999) anima a los psicólogos a darle más importancia a la cultura, ya que "la cultura es un conjunto de sistemas de significado compartido que están inmersos en artefactos y prácticas que conforman un medio para el desarrollo humano". Este tipo de perspectivas culturales deben incorporarse en las teorías psicológicas.

Miller señala investigaciones recientes sobre el autoconcepto para ilustrar los beneficios de la incorporación de la cultura en las teorías psicológicas. Las teorías tradicionales del autoconcepto se basan en la cultura estadounidense y de Europa occidental; el "sí mismo" es un concepto individualista; las personas son independientes unas de otras y el mejoramiento de uno mismo proviene de los logros individuales. Kitayama y sus colaboradores tienen una perspectiva cultural: en contraste con el significado del sí mismo en Estados Unidos, en otras culturas éste es un concepto colectivo en el que la autoestima se deriva de las relaciones con los demás. Los japoneses se enfocan a la autocrítica, mientras que los estadounidenses lo hacen a la autosuperación; por ende, son actividades muy diferentes las que contribuyen a un autoconcepto positivo en cada una de las dos culturas (Kitayama, Markus, Matsumoto y Norasakkunkit, 1997).

También es importante ser conscientes de las formas en que las definiciones operacionales de los constructos que estudiamos, se fundamentan en un significado cultural particular. Una medida de la autoestima que es apropiada para una cultura individualista, probablemente no lo sea en una cultura colectiva, en la cual podría producir resultados erróneos.

GENERALIZACIÓN A OTROS EXPERIMENTADORES

El individuo que lleva cabo el experimento es fuente de otro problema en la generalización. En la mayoría de las investigaciones se utiliza un solo experimentador, y en pocas ocasiones se pone atención a las características personales de éste (McGuigan, 1963). La meta principal es asegurarnos de mantener constante cualquier influencia que el experimentador tenga sobre los sujetos en el transcurso del experimento. Sin embargo, siempre existe la posibilidad de que los resultados puedan generalizarse únicamente a cierto tipo de experimentadores.

Kintz y sus colaboradores han estudiado algunas de las características importantes de los experimentadores (Kintz, Delprato, Mettee, Persons y Schappe, 1965). Estas incluyen la personalidad y el género del experimentador así como su experiencia en este papel. Es muy probable que un experimentador cálido y amistoso produzca resultados distintos a un experimentador frío y poco amistoso. También es posible que los participantes se comporten de forma diferente con un experimentador que con una experimentadora. ¡Se ha demostrado que

los conejos aprenden más rápido cuando son entrenados por experimentadores con experiencia (Brogden, 1962)! La influencia del experimentador también puede depender de las características de los participantes. Por ejemplo, aparentemente los participantes se desempeñan mejor cuando un experimentador del sexo opuesto aplica las pruebas (Stevenson y Alien, 1964).

Una solución al problema de la generalización a otros experimentadores, es la participación de dos o más investigadores, preferentemente hombre y mujer, al realizar el estudio. Un ejemplo adecuado del uso de múltiples experimentadores es el estudio de Rubin (1975), quien envió a varios experimentadores, hombres y mujeres, al aeropuerto de Boston, para investigar qué tanto revela la gente de sí misma. Los experimentadores revelaron distintos tipos de información de sí mismos a viajeros masculinos y femeninos, y registraron la información que a su vez revelaron los pasajeros.

PRETEST Y GENERALIZACIÓN

Con frecuencia, los investigadores se enfrentan a la decisión de aplicar un pretest, ya que intuitivamente, parece ser una buena idea hacerlo. El investigador puede asegurarse de que los grupos sean equivalentes en el pretest y es más satisfactorio ver que cambiaron los puntajes de los individuos, que sólo revisar medias grupales en un postest. Sin embargo, los pretest pueden limitar la posibilidad de generalizar los resultados a poblaciones que no los respondieron. En la vida real, es raro que una persona responda un pretest; por ejemplo, no se miden las actitudes antes de escuchar un discurso político o ver un anuncio véase Lana, 1969).

Una razón importante para el uso de un pretest es que permite que el investigador evalúe los efectos de mortandad, cuando existen posibilidades de que algunos participantes abandonen el experimento. Si usted aplica un pretest, puede determinar si las personas que abandonaron el estudio son diferentes de aquellas que lo completaron. En el capítulo 8 estudiamos que podemos utilizar el diseño de cuatro grupos de Solomon (Solomon, 1949) en situaciones en que es deseable la aplicación de un pretest, pero existe la preocupación de su posible impacto. En el diseño de cuatro grupos de Solomon, la mitad de los participantes resuelve el pretest y la otra mitad únicamente el postest. En otras palabras, el mismo experimento se lleva a cabo con y sin el pretest. Los efectos de la mortandad pueden evaluarse en las condiciones de pretest, y el investigador también puede examinar si existe una interacción entre la variable independiente y la variable aplicación o no del test; es decir, ¿difieren las puntuaciones de la variable dependiente en el postest, si se aplicó el pretest? En ocasiones, los investigadores descubren que no es posible realizar el estudio con los cuatro grupos en un solo experimento. En este caso, el primer estudio puede incluir el pretest; el estudio puede replicarse posteriormente sin el pretest.

GENERALIZACIÓN A PARTIR DE AMBIENTES DE LABORATORIO

La investigación que se realiza en un ambiente de laboratorio tiene la ventaja de permitir que el experimentador estudie el impacto de variables independientes bajo condiciones muy controladas. Sin embargo, surge la duda de si la artificia-lidad del ambiente de laboratorio limita la posibilidad de generalizar, lo que se observa en el laboratorio, a un ambiente de la vida real. Los experimentos de campo, estudiados en el capítulo 4, representan un método para contrarrestar la artificialidad del laboratorio. En un experimento de campo, el investigador manipula la variable independiente en un ambiente natural: una fábrica, una escuela o la esquina de una calle, por ejemplo.

Realismo mundano y experimental

Aronson, Brewer y Carlsmith (1985) señalan que incluso los experimentos de campo pueden ser falsos y artificiales, con poca relevancia para la vida real. Ellos proponen una mejor distinción: el realismo mundano contra el realismo experimental. El **realismo mundano** se refiere a la similitud que tiene el experimento con los eventos que ocurren en la vida real. El **realismo experimental** se refiere al impacto que tiene el experimento sobre los participantes, si los involucra y logra que tomen en serio el estudio. Aronson y sus colaboradores señalan que el realismo experimental y el mundano constituyen dimensiones separadas e independientes. Un experimento puede incluir bastante realismo mundano (y, por lo tanto, ser muy similar a la vida real), pero ser totalmente aburrido y no involucra a los participantes, careciendo así de realismo experimental. Considere dos ejemplos:

> *Ejemplo 1:* en un estudio sobre la conducta de ayuda, un experimentador deja caer algunos lápices de un edificio de un *campus* universitario, y espera a que alguien se detenga y lo ayude a recogerlos. La variable dependiente es la ayuda brindada por los observadores del "accidente". Este estudio tendría mucho realismo mundano, pero poco de tipo experimental.
>
> *Ejemplo 2:* en otro experimento sobre la conducta de ayuda, los participantes se sientan en cubículos individuales que están interconectados por un sistema de intercomunicación. Durante el transcurso del estudio, el individuo escucha que otro (supuesto) participante tiene una crisis epiléptica y debe decidir si ayuda a la persona (Darley y Latané, 1968). Este estudio posee relativamente poco realismo mundano, pero un gran realismo experimental.

Otros procedimientos experimentales, tanto en ambientes de laboratorio como de campo, pueden tener altos o bajos niveles de realismo mundano y experimental. Los experimentos con poco realismo mundano (con poca similitud con eventos o tareas reales) y poco realismo experimental (que no logran involucrar o despertar el interés de los participantes) no suelen producir resultados valiosos. Sin embargo, lo más importante es que los experimentos de labora-

torio no son automáticamente artificiales; aun cuando un experimento carezca de realismo mundano, puede ser muy realista en términos experimentales.

Beneficios recíprocos de la investigación de laboratorio y de campo

Aronson y sus colaboradores (1985), también señalan que es inadecuado considerar a los experimentos de laboratorio o de campo de manera aislada. Realizar investigaciones en ambientes de laboratorio y de campo proporciona la mayor oportunidad para lograr una mayor comprensión del comportamiento. Recuerde el estudio de Langer y Rodin (1976), sobre los efectos de otorgar mayor control en las decisiones de su vida a los residentes de asilos (véase el capítulo 4). Este experimento de campo formó parte de una tradición de investigación que incluye estudios de laboratorio sobre el control de estresores por medio del uso de sujetos animales y humanos, de los efectos del estrés en ambientes naturales y de experimentos de campo sobre los efectos generales de la percepción del control. Anderson, Lindsay y Bushman (1999) se preguntaron si los experimentos de laboratorio y de campo que estudian las mismas variables producen, de hecho, los mismos resultados. Para responder esta pregunta, analizaron 38 pares de estudios de investigaciones de laboratorio, con su contraparte experimental de campo. Los estudios se obtuvieron de una gran variedad de áreas de investigación, incluyendo agresión, conducta de ayuda, memoria, estilo de liderazgo y depresión. Los resultados de los experimentos de laboratorio y de campo eran bastante similares; la magnitud del efecto de la variable independiente sobre la variable dependiente fue muy similar en los dos tipos de estudios. Así, aun cuan-do los experimentos de laboratorio y de campo se llevan a cabo en ambientes distintos, los resultados son complementarios, más que contradictorios.

LA IMPORTANCIA DE LAS REPLICACIONES

Se ha destacado el uso de las **replicaciones** para resolver cualquier problema de generalización que se presente en un estudio. Existen dos tipos de replicaciones: las directas y las conceptuales.

Replicaciones directas

Una **replicación directa** es un intento por repetir de forma precisa los procedimientos de un estudio, para ver si se obtienen los mismos resultados. El investigador que obtiene un hallazgo inesperado, con frecuencia intentará hacer una replicación para asegurarse de que el hallazgo es confiable. Si usted está empezando a trabajar en un problema, puede tratar de replicar un estudio crucial para asegurarse de que comprende los procedimientos y que puede obtener los mismos resultados. Con frecuencia, se realizan replicaciones directas cuando un investigador se basa en los hallazgos de un estudio previo. Por ejemplo,

suponga que se siente intrigado por la investigación de Singh (1993) sobre la proporción de la cintura y la cadera, mencionada en el capítulo 5. Singh reporta que los hombres consideran que las mujeres con una proporción de 0.70 son las más atractivas. En su investigación, usted podría repetir los procedimientos empleados en el experimento original y ampliar la investigación. Por ejemplo, podría estudiar este fenómeno con hombres similares a los de la muestra original, así como también con hombres de diferentes culturas o grupos de edad. Cuando replica hallazgos originales de investigación, por medio de procedimientos muy similares, aumenta su confianza en la generalización de los primeros hallazgos.

En ocasiones, un investigador es incapaz de replicar un hallazgo previo. Sin embargo, un fracaso en la replicación no revela mucho; es poco realista suponer, con base en un solo fracaso, que la investigación previa no es válida. Los fracasos en la replicación comparten los mismos problemas de los resultados no significativos que estudiamos en el capítulo 13. Un fracaso podría significar que los resultados originales no son válidos, pero también podría implicar que el intento de replicación fracasó. Por ejemplo, si la replicación se basa en el procedimiento, tal como se reportó en un artículo científico, es posible que el artículo haya omitido un aspecto importante del procedimiento. Por esta razón, es adecuado establecer contacto con el investigador para obtener información detallada sobre todos los materiales utilizados en el estudio.

El "efecto Mozart" ofrece un ejemplo interesante de la importancia de las replicaciones. En el estudio original, realizado por Rauscher, Shaw y Ky (1993), estudiantes universitarios escucharon una sonata de Mozart durante 10 minu-tos. Después, los estudiantes mostraron un mejor desempeño en una prueba de razonamiento espacial, tomada de la Escala de Inteligencia Stanford-Binet. Posteriormente, Rauscher, Shaw y Ky (1995) replicaron el efecto utilizando una prueba de habilidad espacial diferente, la tarea de doblar y cortar papel de Stanford-Binet. A pesar del hecho de que el efecto fue temporal, de alrededor de 10 minutos, estos hallazgos recibieron mucha atención de la prensa, y las perso-nas rápidamente los generalizaron a la posibilidad de aumentar la inteligencia de los niños con las sonatas de Mozart. De hecho, un gobernador estatal em-pezó a producir discos compactos de Mozart y a distribuirlos en maternidades, y algunos empresarios empezaron a vender música de Mozart para los padres a través de Internet. Sin embargo, durante los siguientes años hubo muchos fra-casos en los intentos de replicar el efecto Mozart (véase Steele, Bass y Crook, 1999). Antes señalamos que los fracasos en las replicaciones pueden ocurrir debido a que no se utilizan las condiciones exactas para producir el efecto. En este caso, Rauscher y Shaw (1998) respondieron a las muchas replicaciones fa-llidas, describiendo con precisión las condiciones necesarias para producir el efecto Mozart. Sin embargo, Steele y sus colaboradores (1999) y McCutcheon (2000) fueron incapaces de lograr el efecto, aun cuando siguieron las recomen-daciones de Rauscher y Shaw. Las investigaciones sobre este efecto continúan. Algunos resultados recientes sugieren que el efecto está limitado a la música que incrementa la activación; es la activación la que puede provocar un mejor

desempeño, después de la exposición a la música (Thompson, Schellenberg y Husain, 2001).

Un solo fracaso en la replicación no es causa suficiente para descartar los hallazgos de la investigación original. Como sucedió en el caso del efecto Mozart, los intentos de replicación no ocurren de forma aislada, sino que muchos investigadores intentan hacerlo. Fracasos repetidos en las replicaciones pueden conducir a la conclusión de que los resultados originales fueron fallidos, es decir, que se cometió un error tipo I. Otra posibilidad es que la investigación de-muestre que los resultados pueden obtenerse únicamente bajo ciertas circuns-tancias. En pocos casos, puede ocurrir que, de alguna forma, el investigador original malinterpretó los resultados (véase en el capítulo 3 la explicación del papel de las replicaciones para la detección de casos de fraude).

Replicaciones conceptuales

El uso de distintos procedimientos para replicar los hallazgos de una investigación se *conoce* como **replicación conceptual.** Las replicaciones conceptuales son aún más importantes que las directas para aumentar nuestra comprensión del comportamiento.

En la mayoría de las investigaciones, la meta es descubrir si existe una relación entre variables conceptuales. La música manipulada en el estudio original del efecto Mozart, fue la primera sección de la Sonata para Dos Pianos en Re Mayor (k448) de Mozart. Ésta es una definición operacional específica para los propósitos del estudio de los efectos de la música sobre un tipo de desempeño espacial. De la misma manera, la tarea específica elegida como medida dependiente es una definición operacional de la variable más general del desempeño.

En una replicación conceptual, se manipula la misma variable independiente de diferente manera (muy posiblemente la variable dependiente también se mida en forma distinta). Las replicaciones conceptuales son sumamente importantes para las ciencias sociales, debido a que las manipulaciones y medidas específicas suelen ser definiciones operacionales de variables complejas. Una pregunta crucial sobre la generalización es si la relación ocurre al estudiar otras formas de manipulación o medición de la variable. En ocasiones, la replicación conceptual puede incluir un estímulo alternativo (por ejemplo, otra sonata de Mozart, una obra de un compositor distinto) o una medida dependiente alternativa (por ejemplo, otra tarea de razonamiento espacial). O, como señalamos anteriormente, en ocasiones se estudian las mismas variables en ambientes de laboratorio y del campo. Cuando las replicaciones conceptuales producen resultados similares, se incrementa en forma importante nuestra confianza en la posibilidad de generalizar las relaciones entre las variables.

Esta discusión también debe alertarlo respecto de una forma importante de pensar acerca de los resultados de investigación. Los hallazgos representan relaciones entre variables conceptuales, pero están fundamentados en operaciones específicas. Usted puede leer acerca de los métodos específicos utilizados en un estudio que se llevó a cabo hace 20 años, y preguntarse si la investigación podría

replicarse hoy en día. También podría especular que los métodos utilizados en un estudio son tan raros que nunca podrían generalizarse a otras situaciones. Estos aspectos no son tan graves cuando se ubican dentro del contexto de las replicaciones conceptuales. Es cierto que un método específico, utilizado en un estudio realizado en un momento dado, podría no ser eficaz en la actualidad, dados los cambios en el clima político y cultural actuales. Sin embargo, una replicación conceptual de la manipulación demostraría que la relación entre las variables conceptuales teóricas aún está presente. De forma similar, el enfoque estrecho de un estudio en particular es menos problemático si el hallazgo general se replica con distintos procedimientos.

EVALUACIÓN DE LA GENERALIZACIÓN MEDIANTE REVISIONES DE LA LITERATURA Y META-ANÁLISIS

Por tradición, los investigadores sacan conclusiones sobre la posibilidad de generalizar hallazgos de investigación al realizar revisiones de la literatura. En una revisión de la literatura, el revisor lee varios estudios que abordan un tema particular y después escribe un artículo que resume y evalúa la literatura. El *Manual de Publicaciones de la Asociación Psicológica Americana* brinda la siguiente descripción: "los artículos de revisión, incluyendo el meta-análisis, son evaluaciones críticas de material previamente publicado. Al organizar, integrar y evaluar material publicado, el autor de un artículo de revisión considera el progreso de la investigación actual, con fines de aclarar un problema" (APA, 2001, p. 7). La revisión de la literatura proporciona información que: 1) resume los hallazgos, 2) le indica al lector cuáles hallazgos tienen un sustento fuerte y cuáles uno débil, 3) señala hallazgos inconsistentes y áreas que requieren ser investigadas y 4) plantea las direcciones futuras de la investigación.

En ocasiones, las revisiones son narrativas en las que el autor describe hallazgos de investigaciones y llega a conclusiones acerca de la literatura. Las conclusiones de las revisiones narrativas de la literatura se basan en las impresiones subjetivas del revisor. En los últimos años ha surgido otra técnica para comparar un gran número de estudios en un área. Esta técnica se llama **meta-análisis** (Rosenthal, 1991). En un meta-análisis, el investigador combina los resultados reales de diversos estudios. El análisis consiste de un conjunto de procedimientos estadísticos que utilizan la magnitud del efecto para comparar un hallazgo en distintos estudios. En lugar de basarse en los juicios obtenidos en una revisión narrativa de la literatura, se pueden sacar conclusiones estadísticas. Usted no debe preocuparse por los procedimientos estadísticos. Éstos implican el examen de varios aspectos de los resultados de los estudios, incluyendo la magnitud del efecto y el nivel de significancia obtenidos. Lo importante aquí, es que el meta-análisis es un método para determinar la confiabilidad de un hallazgo cuando se examinan los resultados de muchos estudios diferentes.

Un ejemplo de meta-análisis es el estudio realizado por Smith y Glass (1977), sobre la eficacia de técnicas de psicoterapia, reportadas en 375 estudios. Los

investigadores examinaron los efectos reportados de diferentes tipos de terapia (por ejemplo, conductual, psicodinámica y centrada en el cliente) a lo largo de muchos estudios diferentes, con diversas medidas de resultados (por ejemplo, reducción de la ansiedad, autoestima). Después aplicaron técnicas estadísticas adecuadas para combinar y comparar los distintos resultados. Su investigación les permitió concluir que la terapia tiene un efecto benéfico. De hecho, concluyeron que "el paciente típico es 75 por ciento mejor que los individuos que no recibieron tratamiento". Los autores también pudieron hacer afirmaciones generales que se aplican a distintos tipos de terapia. Por ejemplo, algunas terapias, como la desensibilización sistemática, producen grandes cambios en el comportamiento, en comparación con otras; sin embargo, en general, no hubo diferencias entre las terapias de tipo "conductual" y las "tradicionales".

La información obtenida de un meta-análisis como el realizado por Smith y Glass, suele brindar gran cantidad de información. En una revisión tradicional de la literatura, sería muy difícil llegar al tipo de conclusión general que se logra con el meta-análisis. Para cualquiera sería muy difícil integrar los resultados de tantos estudios con diseños experimentales, sujetos y medidas distintos. De hecho, si usted leyera todos los estudios y alguien le planteara la sencilla pregunta: "¿funciona la psicoterapia?", tal vez tardaría todo el día informando a la persona acerca de los estudios específicos y las complejidades que encontró al leer la literatura. Este sería el resultado de un exceso de información y del hecho de que es muy difícil integrar información que proviene de diversas fuentes.

Un estudio interesante, llevado a cabo por Cooper y Rosenthal (1980), demostró que los investigadores son propensos a sacar conclusiones más fuertes acerca de un conjunto de hallazgos, cuando utilizan el meta-análisis, que cuando utilizan juicios tradicionales. En este estudio, los investigadores leyeron únicamente siete artículos sobre las diferencias por género en la persistencia al realizar una tarea; aun así, los investigadores que utilizaron el meta-análisis tendieron a concluir que las mujeres son más persistentes que los hombres.

El meta-análisis se utiliza cada vez más cuando se examinan relaciones entre variables. Por ejemplo, Saks y Marti (1997) examinaron muchos estudios que comparan los resultados de jurados conformados de 6 y 12 personas: algunos de los estudios implicaban experimentos que manipularon la variable con jurados simulados; otros utilizaron métodos no experimentales para comparar los resultados de jurados reales, en estados que permiten jurados conformados de 6 personas. Saks y Marti categorizaron los estudios según los resultados de la variable dependiente, como el tiempo de deliberación, la representación de minorías, los jurados divididos y la evocación de testimonios. Los autores encontraron que los jurados más numerosos tienen una mayor representación de grupos minoritarios y mayor división de las decisiones. Con base en una cantidad pequeña de estudios, parece que un jurado más numeroso delibera durante más tiempo y recuerda los testimonios con mayor precisión. Una de las características más importantes de los estudios de meta-análisis es su enfoque en la magnitud del efecto. Esto puede llevarnos a la conclusión de que algunos hallazgos son mucho más débiles de lo que antes se pensaba. Por ejemplo, mu-

chos estudios sobre los efectos del atractivo físico apoyan el estereotipo de que "lo que es bello es bueno"; es decir, que las personas físicamente atractivas reciben juicios más favorables que las personas menos atractivas. El hecho de que muchos estudios apoyen esta conclusión, puede llevarnos a sobreestimar su importancia: los meta-análisis de los estudios del atractivo físico concluyen que el efecto del atractivo físico sobre muchas variables, incluyendo las decisiones de jurados, es muy pequeño (Eagly, Ashmore, Makhijani y Longo, 1991; Mazzella y Feingold, 1994).

Las revisiones narrativas y los meta-análisis brindan información valiosa y, de hecho, son complementarios. Un meta-análisis permite llegar a conclusiones estadísticas y cuantitativas, mientras que la revisión narrativa identifica tendencias de la literatura y la tendencia de estudios futuros, es decir, es un método más cualitativo. Un estudio realizado por Bushman y Wells (2001), señala una forma interesante en la que el conocimiento derivado del meta-análisis puede mejorar la forma en que interpretamos la información de las revisiones de la literatura. En su estudio, los revisores eran estudiantes de licenciatura a quienes se les dio el título e información acerca de los hallazgos de 20 estudios respecto a los efectos de la similitud de las actitudes sobre la atracción. Algunas veces el título era relevante con respecto a los hallazgos ("Dios los hace y ellos se juntan"), y otras no lo era ("investigación que estudia quién gusta de quién"). Obviamente, los títulos relevantes son más fáciles de recordar. Cuando se les pidió que sacaran conclusiones acerca de los hallazgos, los revisores inexpertos, que carecían de conocimientos acerca del meta-análisis, sobreestimaron la magnitud de la relación similitud-atracción, cuando se les dieron títulos relevantes. A otros revisores se les dio un breve entrenamiento en meta-análisis; estos revisores llegaron a conclusiones precisas acerca de los hallazgos; no se vieron influidos por el título del artículo. Por lo tanto, aún sin realizar un meta-análisis, los conocimientos sobre este método pueden beneficiar a quienes revisen hallazgos de investigación.

USO DE LA INVESTIGACIÓN PARA MEJORAR LAS CONDICIONES DE VIDA

En un discurso presidencial de la Asociación Psicológica Americana, George Miller (1969) consideró a la "psicología como un medio de promoción del bienestar humano", y habló de "regalar la psicología". Miller se refería al amplio tema de la generalización; tomar lo que sabemos sobre el comportamiento humano y permitir que sea aplicado por muchas personas en todas las áreas de la vida cotidiana. Quizás la evidencia más fuerte de la generalización de los resultados de investigación consista en ver lo que ha sucedido con la psicología desde que Miller nos invitó a regalarla. El impacto de la investigación psicológica puede observarse en los campos del cuidado de la salud (programas diseñados para promover comportamientos saludables con relación al estrés, las enfermedades cardiacas y las enfermedades de transmisión sexual); del ambien-

te (al diseñar espacios de trabajo que promuevan la satisfacción y productividad del trabajador; al proporcionar a empresas de energía fórmulas que promueven la conservación); de las leyes (al proporcionar datos sobre los efectos de jurados de 6 y 12 personas, al mostrar cómo el personal que aplica las leyes puede incrementar la precisión de la identificación de testigos); y de la educación (al proporcionar métodos para mejorar el desempeño académico o reducir el conflicto entre distintos grupos étnicos). Además, los psicólogos están utilizando la Internet para brindar información sobre la paternidad, la educación, la salud mental y muchos otros temas. Por ejemplo, las páginas de la American Psycho-logical Association y la American Psychological Society (http://www.apa.org; http://www.psychologicalscience.org), el Center for Mental Health Services ¡http://www.mentalhealth.org) y muchos psicólogos que en lo individual comparten su experiencia con el público. Hemos comentado sólo algunas de las formas en que la investigación básica se ha aplicado con el fin de mejorar las condiciones de vida. A pesar de todos los problemas potenciales de la generalización de los hallazgos de investigación que se señalaron en este capítulo, las evidencias sugieren que podemos generalizar nuestros resultados a muchos aspectos de nuestras vidas.

Términos de estudio

Diseño de cuatro grupos de Solomon
Meta-análisis
Realismo experimental
Realismo mundano
Replicación
Replicación conceptual
Replicación directa
Revisión de la literatura
Validez externa

Preguntas de repaso

1. ¿Por qué debe preocuparse un investigador por generalizar sus hallazgos a otras poblaciones de sujetos? ¿Cuáles son algunos de los problemas de la generalización a poblaciones de sujetos, que debe enfrentar un investigador?
2. ¿Cuál es el origen del problema de la generalización a otros experimentadores? ¿Cómo puede resolverse este problema?
3. ¿Por qué la aplicación de pretest constituye un problema de generalización importante? Describa cómo el uso de pretest puede afectar la generalización.
4. Establezca la diferencia entre el realismo mundano y el experimental.
5. Establezca la diferencia entre una replicación directa y una replicación conceptual. ¿Cuál es el valor de una replicación conceptual?
6. ¿Qué es un meta-análisis?

Actividades

1. Es fácil reunir datos de experimentos y encuestas en Internet. Cualquier persona del mundo con una computadora que esté conectada a Internet puede tener acceso a la computadora de un investigador y tomar parte en el estudio. Participe en un estudio de investigación psicológica de Internet. ¿Qué aspectos de la generalización pueden surgir cuando se interpretan los resultados de este tipo de estudios? ¿El hecho de utilizar una computadora hace que la investigación tenga menos posibilidades de generalizarse que la investigación tradicional? ¿O el hecho de que puedan participar personas de todo el mundo la hace más generalizable? ¿Podría responder empíricamente esta pregunta?
2. Utilice *PsycINFO* para encontrar resúmenes de artículos que incluyan la raza, el grupo étnico, el género o la nacionalidad como variable en los estudios. ¿Qué conclusiones plantean los autores de estos estudios acerca de la generalización?
3. Encuentre un meta-análisis publicado en una revista; dos buenas fuentes son el *Review of Educational Research y* el *Psychological Bulletin.* ¿Qué conclusiones se sacaron del meta-análisis? ¿Cómo se seleccionaron los estudios para el meta-análisis? ¿De qué forma se discutió el concepto de la magnitud del efecto en el meta-análisis?
4. Obtenga un artículo de una revista de psicología social. Después de resumir brevemente el procedimiento experimental, comente el experimento en términos del realismo mundano y del realismo experi-mental.

Apéndice A
Redacción de reportes de investigación[1]

INTRODUCCIÓN

Este apéndice presenta la información que usted necesitará para preparar un reporte escrito de investigación, ya sea para un curso o para su posible publicación en una revista especializada. Asimismo, se incluye un artículo que ilustra las características de estilo de un reporte de investigación. Consideraremos las reglas específicas que deben seguirse para organizar y presentar resultados de investigación. Estas reglas son muy convenientes para el escritor y para el lector, debido a que proporcionan estructura al reporte y un método uniforme de presentación, lo que facilita al lector la comprensión y evaluación del documento.

Las reglas específicas varían de una disciplina a otra, es posible que las reglas para presentar los resultados en psicología no puedan aplicarse a la investigación sociológica, por ejemplo. Además, las reglas varían si se está preparando el reporte para una clase, una tesis o para una revista científica. Por fortuna, las variaciones suelen ser pequeñas y las reglas generales de presentación son bastante similares en las distintas disciplinas y situaciones.

El formato para redactar reportes de investigación que presentamos aquí, se basa en el *Manual de Publicaciones de la Asociación Psicológica Americana* (APA, quinta edición, 2001). El estilo de la APA se emplea en muchas revistas científicas de psicología, de desarrollo infantil, de relaciones familiares y de educación. Si usted se interesa por las reglas específicas de una revista en particular, consulte un número reciente. La APA también publicó un libro de trabajo y una guía de entrenamiento del estudiante para el *Manual de Publicaciones* (Gelfand y Walker, 2001), y un programa de cómputo llamado *APA-Style Helper*, que funciona con casi todos los procesadores de textos, para asegurarse de que el trabajo se ajuste a las reglas de estilo de la APA. Si a usted le interesan éstos y

[1] **Nota de los revisores técnicos.** En razón de que muchas reglas de estilo difieren entre los idiomas inglés y español, en este apéndice se han conservado mayormente los lincamientos que son aplicables en ambos idiomas. Se le sugiere que cuando elabore un trabajo para publicación analice las normas específicas de la revista a la que lo piensa enviar. Si prepara su trabajo en inglés, puede consultar el *Manual de Publicaciones* de la APA; si su trabajo es en español, existe una versión del mismo manual, publicada por *El Manual Moderno.*

otros productos de la APA, los sitios de Internet http://www.apa.org y http://www.apastyle.org, indican la forma más sencilla de obtenerlos. Otras fuentes útiles para la preparación de trabajos son las obras breves de Rosnow y Rosnow (2002) y Sternberg (1993). Kazdin (1995) y Bem (2002) también ofrecen lineamientos excelentes para preparar reportes de investigación para ser publicados. El manual de la APA se guía por los principios de "especificidad y sensibilidad". En primer lugar, los trabajos deben escribirse con un nivel de especificidad y detalle que permita que otros autores repliquen la investigación. En segundo lugar, los trabajos deben estar libres de lenguaje inapropiado que pueda ser interpretado de manera prejuiciosa e insensible. El manual también incluye lineamientos para la preparación del manuscrito, que aprovechan los beneficios de las características de los programas de procesadores de textos. A lo largo de este apéndice, se incluyen ejemplos de cómo aparecería la información, utilizando máquina de escribir, para aclarar las reglas del espaciamiento y otro tipo de lineamientos que también se utilizan en el manual de la APA.

ESTILO DE REDACCIÓN

El estilo de redacción es importante, no importa cuál sea el formato que utilice para preparar el reporte. Un reporte con una redacción pobre, que es difícil de entender, no tiene valor alguno (¡y casi con certeza le proporcionará una baja calificación!). Además, un trabajo con calidad debe ser escrito con limpieza y libre de errores de ortografía y de tipografía.

Claridad

La claridad de la redacción es esencial. Sea preciso y claro al presentar sus ideas, y piense en los lectores a quienes va dirigido. Es útil dirigir su trabajo a lectores que no estén familiarizados con su tema general ni con los métodos que utilizó para estudiarlo. Elimine los términos que la mayoría de los lectores no comprenderían. En ocasiones, los investigadores crean notaciones abreviadas para referirse a una variable o un procedimiento específico. Este tipo de abreviaturas puede ser conveniente cuando el investigador se comunica con otras personas que están involucradas directamente en el proyecto de investigación, pero son confusas para los lectores en general. Sin embargo, usted debe asumir que el lector está familiarizado, hasta cierto punto, con la estadística y con la prueba de hipótesis. Generalmente, los resultados estadísticos pueden presentarse sin definir términos como *media*, *desviación estándar* o *significación.* Sin embargo, éstos son únicamente lineamientos generales. Rosnow y Rosnow (2002) señalan que cuando el trabajo está dirigido al maestro, ¡se debe poner mucha atención a las expectativas que éste tiene respecto al reporte!

Todo el trabajo debe poseer una estructura coherente. Las ideas deben presentarse en un orden lógico para facilitar su comprensión. Si usted escribe el reporte para alguien que por primera vez tiene contacto con sus ideas y con

hallazgos de investigación, deberá poner más atención en comunicarse con gran claridad.

Un método para producir un reporte más organizado es el uso de un guion. Muchos escritores planean un reporte plasmando sus ideas en forma de bosquejo, el cual después sirve como una guía. Este método obliga a que los escritores elaboren una estructura lógica antes de redactar el trabajo. Otros escritores prefieren utilizar un método menos estructurado para realizar el primer borrador. Después tratan de estructurar lo que ya está escrito. Si el trabajo no produce un guión coherente, debe de organizarse mejor. Los programas de procesadores de textos generalmente incluyen una herramienta de guión que ayuda a organizar su reporte; localice esta herramienta en el menú de ayuda del programa.

Los párrafos deben estar bien organizados. Es bueno que un párrafo tenga un enunciado del tópico o idea central. Los otros enunciados del párrafo deben estar relacionados con el enunciado del tópico y deben desarrollar la idea al elaborarla, ampliarla o apoyarla. Además, evite párrafos de un solo enunciado. Si encuentra párrafos de este tipo en su trabajo, extiéndalos, incluya la idea en otro párrafo o elimine el concepto. Después de completar el primer borrador, es conveniente dejar pasar uno o dos días antes de retomarlo. Lea cuidadosamente el trabajo, poniendo especial atención a la gramática y la ortografía. Aquí describimos algunos aspectos gramaticales; también puede usar un programa de procesador de textos para verificar la gramática y la ortografía. Después de hacer cambios y correcciones, puede recibir retroalimentación de otros. Pida a una o más personas que lea su reporte de manera crítica y que le sugieran formas de mejorarlo. Esté usted preparado para escribir varios borradores antes de lograr un producto final satisfactorio.

Reconocimiento del trabajo de otros

Es sumamente importante separar con claridad sus propias palabras e ideas de aquellas obtenidas de otras fuentes. Si utiliza un párrafo de un artículo o libro, asegúrese de presentarlo como una cita textual. No hay nada malo en citar a otro autor, siempre y cuando reconozca la fuente. Nunca presente las ideas de otra persona como suyas, ya que incurriría en un plagio y esto es imperdonable. En ocasiones, los escritores se ven tentados a llenar su trabajo con citas textuales de otras fuentes o a incluir citas muy extensas de otro trabajo (por ejemplo, varios párrafos). Esta práctica produce distracción y es contraproducente. Sea directo y utilice sus propias descripciones e interpretaciones, mientras dé reconocimiento a sus fuentes. Si tiene alguna duda acerca de la forma apropiada de incluir material de otros artículos en su trabajo, consulte a su maestro.

Voz activa y voz pasiva

Muchos escritores utilizan en exceso la voz pasiva en sus reportes, quizás debido a que consideran que la voz pasiva le da una apariencia más "científica" a su escrito. Considere los siguientes enunciados:

Fue descubierto por Yee y Johnson (1996) que los adolescentes prefieren.. .

A los participantes les fue administrada la prueba después de un periodo de descanso de 10 minutos.

A los participantes se les leyeron las instrucciones del experimento.

Ahora, trate de escribir los mismos enunciados en una voz activa. Por ejemplo:

Yee y Johnson (1996) descubrieron que los adolescentes prefieren. . .

Los participantes respondieron la prueba después de un periodo de descanso de 10 minutos.

El experimentador leyó las instrucciones a los participantes.

La prosa que suena forzada al redactarse en voz pasiva, es mucho más directa cuando se expresa en voz activa.

A veces, los autores se refieren a sí mismos en tercera persona, de modo que pueden decir "el experimentador distribuyó los cuestionarios" en lugar de "*yo* distribuí los cuestionarios"; la primera es una voz activa, pero introduce ambigüedad. No queda claro si "el experimentador" es el propio autor o alguien más. Cuando los autores se refieren a sí mismos en un reporte, el estilo de la APA sugiere el uso de los pronombres "yo" y "nosotros".

Lenguaje prejuiciado

El estilo de la APA se guía por los principios de especificidad y sensibilidad. El principio de especificidad conlleva la recomendación del uso del término *participantes* para distinguir de manera específica entre sujetos humanos y animales. Además, es apropiado describir a los participantes como encuestados, niños, pacientes, clientes y así por el estilo, si estos términos describen con precisión a las personas que participaron en el estudio.

Sea sensible a la posibilidad de que su escrito contenga prejuicios, no intencionales, con respecto al género, la orientación sexual y el grupo étnico o racial. Como principio general, sea lo más específico posible cuando se refiera a grupos de personas. Por ejemplo, referirse a los participantes de su estudio como "coreanos-americanos y vietnamitas-americanos" es más específico y preciso que describirlos como "asiáticos". Asimismo, sea sensible al uso de etiquetas que podrían ser ofensivas para los miembros de ciertos grupos. En la práctica, esto significa que es mejor referirse las personas utilizando los términos que ellas prefieren. Además, evite etiquetas implícitas al decir "la muestra de lesbianas, en contraste con la muestra de mujeres normales" o "probamos grupos de individuos autistas y normales". Esta última frase podría estar mejor escrita, "probamos personas con autismo y sin autismo".

El manual de la APA incluye numerosos ejemplos de formas de ser sensible al género, la identidad racial y étnica, la edad, la orientación sexual y las disca-

pacidades. El *término género* se refiere a los hombres y las mujeres como grupos sociales; así, género es la palabra adecuada en frases como "diferencias por género en el salario promedio". El término *sexo* se refiere a los aspectos biológicos de los hombres y de las mujeres; por ejemplo, "fantasías sexuales" o "diferencias sexuales en el tamaño de ciertas estructuras cerebrales". El uso de los pronombres de género puede ser problemático. No utilice *él, hombre, del hombre,* etc., cuando se refiera a hombres y mujeres. Para evitar prejuicios lingüísticos, con frecuencia es posible reescribir los enunciados u omitir pronombres específicos. Por ejemplo, "al trabajador se le paga de acuerdo a su productividad" puede cambiarse por "al personal se le paga de acuerdo a su productividad" o "a los trabajadores se les paga de acuerdo a su productividad". En el primer caso, se cambió el término "al trabajador", que hace referencia sólo a hombres, por el término "al personal", que incluye a hombres y mujeres. *No* intente evitar lenguaje sexista al sustituir simplemente *él-ella* siempre que parezca ser conveniente.

El manual de la APA incluye una gran cantidad de información y diversos ejemplos que promueven la sensibilidad en los reportes escritos. El mejor consejo es revisar los trabajos, para encontrar posibles problemas, al menos una vez antes de redactar el escrito final. Si tiene dudas respecto al lenguaje apropiado, consulte con otros colegas, cuya opinión usted respete.

Mecanografía y procesadores de textos

Eventualmente, usted tendrá que preparar una copia mecanografiada de su trabajo. De acuerdo con el estilo de la APA, el reporte debe ir *completamente a doble espacio.* Los márgenes del texto deben ser de *al menos 2.5 cm,* a los cuatro lados de la hoja. Los encabezados de las páginas, la información que aparece al inicio de cada hoja, incluyendo el número de página, se colocan aproximadamente a media pulgada de la orilla de la hoja. Todas las páginas se numeran, con excepción de aquellas que incluyen figuras al final del trabajo. Los párrafos deben tener una sangría de entre 5 y 7 espacios o media pulgada (utilice el tabulador, no la barra espaciadora). En el estilo de redacción propuesto por la APA no se utilizan contracciones. Nunca se deben utilizar guiones al final de un renglón, sino que los renglones deben ser un poco más cortos o más largos para evitar partir una palabra.

Es muy común que los escritores inserten dos espacios después de cada enunciado; sin embargo, según los lincamientos de la APA, solamente debe colocarse un espacio entre enunciados. Los dos espacios eran adecuados para las máquinas de escribir, pero un espacio es más atractivo y fácil de leer cuando se emplea un procesador de textos con fuentes modernas (consejo: si tiene problemas para recordar esta regla, utilice la herramienta de "reemplazar" de su procesador de textos para cambiar los dos espacios a uno).

La mayoría de los estudiantes emplea procesadores de texto para elaborar sus trabajos. Por medio de la computadora, usted puede mejorar su escrito y entregar un trabajo de gran calidad con mayor facilidad. Las correcciones, revisiones, cambios y eliminaciones de texto, así como la preparación de varios borra-

dores, son fáciles porque no tiene que reescribir el trabajo cada vez que desea hacer algún cambio. Asegúrese de sacar ventaja de las herramientas de su procesador de textos. Puede colocar encabezados y números de página de forma automática en la parte superior de cada página, insertar pies o notas de página, centrar el texto, verificar la gramática y la ortografía, etcétera.

Establezca la alineación del texto en el margen izquierdo. La fuente debe tener un tamaño de 12 puntos en todo el reporte. Utilice una fuente serif en todo el texto y en las tablas. Por lo general, la fuente serif debe ser Times Román. Sin embargo, las figuras deben elaborarse con una fuente sans serif, ya sea Arial o Helvética. Las fuentes serif tienen pequeñas líneas al final de los trazos que conforman las letras; sans serif significa literalmente "sin seriP, por lo que no tiene líneas serif He aquí algunos ejemplos:

Éste es texto serif. Éste es
texto san serif.

Cuando requiera el uso de cursivas, emplee la característica "fe" de su procesador de textos. Las cursivas se utilizan para: *a)* título y volumen de revistas periódicas, *b)* título de libros, *c)* encabezados al margen y de párrafos, *d)* la mayoría de los términos estadísticos, *e)* los extremos de una escala, como 1 *(totalmente en desacuerdo)* a 5 *(totalmente de acuerdo) yf)* cuando sea necesario enfatizar una palabra o frase específica, al mencionarla por primera vez en el trabajo. Ponga atención en el uso de cursivas en los ejemplos utilizados en todo este apéndice. Finalmente, existen pocas razones para utilizar las negritas.

Estilo de la APA y formatos de reportes de estudiantes

El formato de la APA está hecho para proporcionar a un tipógrafo un manuscrito, mismo que después prepara el reporte para su publicación en una revista científica; varios de los requisitos de estilo de la APA son de utilidad para el tipógrafo. Sin embargo, cuando prepare un trabajo para un reporte de clase, un proyecto de licenciatura o una tesis, su escrito puede ser el "producto final" para sus lectores. En tales casos, puede ignorar muchos de los aspectos del estilo de la APA, de modo que su trabajo se asemeje lo más posible a un reporte impreso. Por ejemplo, el estilo de la APA requiere que las tablas y las figuras se coloquen al final del trabajo, ya que el tipógrafo las inserta en el cuerpo del reporte para la publicación del artículo. Sin embargo, si su trabajo es la versión final para los lectores, quizás necesite: *a)* colocar las tablas y las figuras en hojas separadas dentro del cuerpo del reporte o *b)* insertar las tablas y las figuras en el cuerpo del texto. A continuación se describen algunas de las diferencias que pueden presentarse entre el reporte de un estudiante y el estilo de la APA. Cuando esté listo para preparar su reporte, asegúrese de verificar los requisitos específicos de su maestro o universidad.

ORGANIZACIÓN DEL REPORTE

Un reporte de investigación está organizado en cinco partes principales: resumen, introducción, método, resultados y discusión. Las referencias deben ordenarse utilizando un formato especial. El reporte también puede incluir tablas y figuras para la presentación de los resultados. Analizaremos las partes del trabajo en el orden recomendado por el estilo de la APA. Remítase al ejemplo de un reporte, al final de este apéndice, conforme lea el siguiente material.

Página del título

La primera página del trabajo incluye el título. Se hace en una hoja separada y se numera como página 1. Observe que en el ejemplo del reporte, la primera página nombra el título; sin embargo, también incluye otros elementos de información importantes.

En la parte superior de la página del título debe aparecer el encabezado, el cual incluye un título breve que generalmente consiste de las primeras dos o tres palabras del título del trabajo, y el número de la página. Estos datos se colocan hacia el margen derecho de la hoja. Si utiliza un procesador de textos, recurra a la herramienta del encabezado de página para crear un encabezado que se imprima aproximadamente a la mitad, entre el texto y la orilla superior de cada hoja, generalmente a media pulgada de la orilla. No intente escribir manualmente el encabezado y el número de la página en cada hoja de su trabajo; revise su programa de procesador de textos y aprenda a crear los encabezados. El encabezado aparece en cada página, con excepción de las páginas de las figuras. El encabezado le recuerda el tema a sus lectores; más importante aún, si las páginas se separan de manera accidental, el lector podrá ordenarlas con facilidad.

En la primera línea de la página se anota el *título principal*. El título principal tiene un significado y propósito muy específicos: se trata de un título abreviado y no debe ser mayor de 50 caracteres (letras, número o espacios). Si el trabajo se publica en una revista científica, el título principal se imprime como encabezado en la parte superior de las hojas para ayudar a que los lectores identifiquen el artículo. El título principal se coloca hacia el margen izquierdo; el título principal del reporte del ejemplo aparece de la siguiente manera:

```
Titulo principal: DESCRIPCIÓN GENÉRICA DE ATLETAS POR PARTE
DE LOS MEDIOS
```

Observe que todas las letras son mayúsculas. El título principal se utiliza para las publicaciones y aparece una sola vez en el trabajo, en la página del título. En contraste, los lectores utilizan el encabezado de la página con el título breve y el número de página, por lo que aparece en cada una de ellas. No confunda el tí-tulo principal con el encabezado de la página.

El resto de la página del título incluye el título, el autor y la afiliación institucional, centrados en la página. El título debe ser corto (por lo general no mayor de 10 a 12 palabras) y debe informar al lector acerca de la naturaleza de la investigación. Una forma adecuada para hacer esto es incluir los nombres de las variables en el título. Por ejemplo, los siguientes títulos son cortos e informativos:

Efectos de la ansiedad en la solución de problemas matemáticos

Memoria para los rostros en adultos jóvenes y mayores

En ocasiones, el uso de dos puntos en el título ayuda a transmitir la naturaleza de la investigación, e incluso añade un poco de "elegancia" al título, como en los siguientes casos

Respuestas cognoscitivas en la persuasión: determinantes afectivos y evaluativos

Comparación de la tortuga y la liebre: diferencias por género y experiencia en tareas dinámicas de razonamiento espacial Otro método para titular un trabajo es plantear la pregunta que aborda la investigación. Por ejemplo,

¿Las recompensas en el salón de clases reducen la motivación intrínseca?

¿Aún existen los estereotipos ocupacionales? Otro aspecto que debe tomarse en cuenta para la elección de un título implica que, en las búsquedas de literatura por computadora, es más fácil encontrar su artículo si el título incluye palabras y frases que las personas utilizan comúnmente al realizar la búsqueda. Esto también se aplica al resumen.

Resumen

El resumen (o *abstract)* es un extracto del reporte de investigación, y generalmente incluye entre 100 y 120 palabras. El propósito del resumen es introducir el artículo, para que los lectores decidan si es relevante para sus propios intereses. El resumen debe proporcionar información suficiente para que el lector decida si le conviene leer el reporte completo; además debe facilitar la comprensión del reporte al leerlo.

A pesar de que el resumen aparece al inicio del reporte, es más fácil escribirlo al finalizar el trabajo. Lea algunos resúmenes y obtendrá ideas adecuadas para condensar un reporte completo en 8 o 10 enunciados llenos de información. Un ejercicio muy informativo consiste en escribir un resumen de un artículo publicado, y después compararlo con el resumen original escrito por los autores.

Los resúmenes generalmente incluyen un enunciado o dos acerca de cada una de las cuatro secciones principales del cuerpo del artículo. En primer lugar,

de la sección de *introducción,* establezca el problema en estudio y la hipótesis principal. En segundo lugar, de la sección del *método,* incluya información sobre las características de los participantes (por ejemplo, número, edad, género y cualquier característica especial) y un breve resumen del procedimiento (por ejemplo, cuestionarios de autorreporte, observación directa, medidas repetidas en varias ocasiones). En tercer lugar, de la sección de *resultados,* describa el patrón de los hallazgos para las variables principales. Esto suele hacerse reportando la dirección de las diferencias, sin incluir valores numéricos. Los lincamientos de la APA recomiendan incluir los niveles de significación estadística, aunque pocos autores lo hacen (Ono, Phillips y Leneman, 1996). Considere los lincamientos establecidos por sus maestros. Finalmente, el resumen debe incluir las implicaciones del estudio, tomadas de la sección de *discusión.* Es preferible incluir comentarios informativos acerca de los hallazgos, y no afirmaciones generales como "las implicaciones del estudio están dirigidas a ..." (Kazdin, 1995).

El resumen se escribe en una hoja separada, numerada como página 2. La palabra "Resumen" se escribe centrada en la parte superior de la página. El resumen siempre se escribe como un solo párrafo, sin sangrías.

Introducción

La sección de introducción se inicia en una nueva página (página 3), con el nombre del reporte escrito en la parte superior. Observe que el nombre del autor no aparece en esta página, lo que permite que un revisor lea el documento sin conocer el nombre del autor. La sección de introducción presenta el problema específico que se estudia, describe la estrategia de investigación y presenta los resultados que fueron predichos. Después de leer la introducción, el lector debe saber por qué el autor decidió realizar la investigación, y cómo la hizo. En general, la introducción parte de teorías y hallazgos generales, hasta llegar a los aspectos específicos del estudio.

La introducción tiene tres componentes, aunque éstos no se presentan como subsecciones formales. Los componentes son: 1) el problema que se estudia, 2) la revisión de la literatura y 3) los razonamientos y las hipótesis del estudio.

La introducción debe iniciar con el planteamiento del problema en estudio. En dos o tres enunciados, el lector debe obtener una apreciación del contexto y la importancia general del tema que se estudia (Bem, 1981; Kazdin, 1995). Vale la pena, si es posible, exponer el problema específico que se está investigando; esto ayuda a que los lectores, incluso aquellos que no están familiarizados con el tema, comprendan y aprecien por qué el autor decidió investigarlo.

Después de la exposición inicial, la introducción proporciona una descripción de las investigaciones y teorías previas. A esto se le llama *revisión de la literatura.* No es necesario realizar una revisión exhaustiva de las teorías e investigaciones anteriores (si existen revisiones relevantes de la literatura del tema, se remite al lector a ellas). En su lugar, se describen únicamente los aspectos teóricos y de investigación que están claramente relacionados con el estudio. Es im-

portante establecer de forma explícita cómo el trabajo previo se asocia de manera lógica con su problema de investigación. Esto indica al lector por qué se llevó a cabo la investigación y muestra la conexión con investigaciones previas. La parte final de la introducción le indica al lector las *razones* del estudio. Aquí se establecen las variables del estudio y los resultados que se esperan; al explicar el origen de las hipótesis, se logra vincular la hipótesis de investigación con los estudios previos y con el diseño de investigación.

Método

La sección del método inicia inmediatamente después del final de la introducción (en la misma página, si el espacio lo permite). Esta sección proporciona al lector información detallada respecto a la realización del estudio. De manera ideal, debe incluir información suficiente para que un lector replique el estudio. Esta sección suele dividirse en varias subsecciones. Tanto el orden de las subsecciones, como su número, varían en los artículos publicados. La decisión acerca de las subsecciones que se incluyen se basa en la complejidad de la investigación. El ejemplo del reporte en este apéndice incluye tres subsecciones: *participantes, diseño y materiales, y procedimiento. A* continuación se explican algunas de las subsecciones más comunes.

Panorama general Si el diseño y los procedimientos experimentales empleados son complejos, debe presentarse un breve panorama general del método para ayudar al lector a comprender la información que sigue.

Participantes Siempre es necesario incluir una subsección acerca de los participantes; se debe describir su número y su naturaleza, así como su edad, género, grupo étnico o social y otras características relevantes. También se describen aspectos especiales como si son primogénitos, hijos adolescentes de alcohólicos, maestros o padres de niños hiperactivos. Se debe establecer explícitamente la forma en que se reclutó a los participantes y los incentivos por su participación utilizados. También se puede incluir el número de individuos en cada condición experimental.

Aparatos Esta subsección puede ser necesaria si se utilizó equipo especial en el experimento. Se debe especificar la marca y el modelo de los aparatos y describir con detalle el equipo. Esto es necesario para replicar el experimento.

Procedimiento Esta subsección le indica al lector la forma exacta en que se llevó a cabo el estudio. Una manera de reportar la información es la descripción, paso a paso, de lo que ocurrió en el experimento. Mantenga la secuencia temporal de los eventos, de modo que el lector sea capaz de visualizar el orden de los sucesos que experimentaron los participantes.

La subsección de procedimiento describe al lector las instrucciones dadas a los participantes, la forma en que se manipularon las variables independientes y

la manera en que se midieron las variables dependientes. También deben describirse los métodos empleados para controlar las variables extrañas, incluyendo los procedimientos de aleatorización, contrabalanceo y medios especiales utilizados para mantener constante una variable en todas las condiciones. Finalmente, es necesario describir el método de informe. Si en el estudio se usó un método no experimental, se deben describir con detalle la forma exacta en que se condujo y las técnicas de medición utilizadas.

Usted debe decidir qué tan detallada será su descripción; es importante que utilice su propio juicio para determinar la importancia de un aspecto específico del procedimiento y la cantidad de detalles necesarios para que el lector comprenda con claridad lo que se hizo. Incluya cualquier detalle que pueda ser importante para replicar el estudio.

Otras subsecciones Si son necesarias, incluya otras subsecciones para lograr una clara presentación del método. Por ejemplo, es probable que se requiera de una subsección que hable de los materiales de prueba, en lugar de una de aparatos. Algunos autores insertan otras secciones que se ajustan a su estudio. Si echa un vistazo a un número reciente de una revista, encontrará que algunos estudios incluyen únicamente dos subsecciones, y que otros utilizan muchas más. Esto refleja la variable complejidad de los estudios y los estilos particulares de redacción de los investigadores.

Resultados

En esta sección se presentan los resultados de la manera más clara posible. La sección de resultados es una descripción directa de sus análisis. Aun cuando es tentador explicar los hallazgos en la sección de resultados, esto se hace en la siguiente sección del trabajo, la sección de discusión.

Asegúrese de hacer explícito el valor de alfa (probabilidad) empleado para la toma de decisiones con respecto a la significación estadística: generalmente se elige el 0.05 o 0.01, y se requiere únicamente un enunciado sencillo como "se utilizó un valor de alfa de 0.05 para los análisis estadísticos".

Presente sus resultados en el mismo orden en que estableció sus predicciones en la sección de introducción del trabajo. Si llevó a cabo una verificación de la manipulación, preséntela antes de describir los resultados principales.

El contenido de la sección de resultados varía de acuerdo al tipo de prueba estadística y a la cantidad de análisis realizados. Sin embargo, toda sección de resultados incluye algunos elementos básicos. Si es pertinente, describa cualquier procedimiento de calificación o codificación de los datos, en su preparación para el análisis. Esto es especialmente importante cuando se codifican datos cualitativos (en ocasiones, la transformación de datos se incluye en una subsección de la sección del método). Establezca qué pruebas estadísticas se realizaron con los datos (prueba í, prueba *F*, correlación, etc.). Justifique la selección de una comparación estadística específica para evaluar su hipótesis. Asegúrese de resumir cada hallazgo en palabras, así como de incluir los resultados de las

pruebas estadísticas en forma de frases estadísticas. La quinta edición del manual de la APA incluye lineamientos para reportar los estadísticos recomendados por un Grupo de Trabajo de la APA sobre Inferencia Estadística (Wilkinson, 1999). Una importante recomendación es reportar los valores exactos de probabilidad proporcionados por los programas de cómputo utilizados para realizar análisis estadísticos. En el pasado, la mayoría de los investigadores reportaban las probabilidades como "menor que" las probabilidades estándar que aparecen en las tablas estadística, por ejemplo, $p < 0.10$, $p < 0.05$ *o* $p < 0.01$. Ahora podemos reportar las probabilidades exactas que tiene la hipótesis nula de ser correcta, por ejemplo, $p = 0.085$, $p = 0.031$ o $p = 0.004$. Este cambio permite que los lectores apliquen sus propios estándares de significación estadística al evaluar el estudio.

Una segunda recomendación es reportar la magnitud del efecto. El manual reconoce que actualmente existen muchos indicadores de la magnitud del efecto, asociados con distintos procedimientos estadísticos; lo importante es incluir una magnitud del efecto en el artículo publicado.

Los resultados deben redactarse con enunciados sencillos. Por ejemplo, los resultados del experimento sobre el modelamiento y la agresión, descritos en el capítulo 12, podrían expresarse de la siguiente manera:

> Como se predijo, los niños que observaron al modelo agresivo se mostraron significativamente más agresivos que los niños en la condición sin modelo, $t(18) = 4.025$, $p = 0.001$. La puntuación media de la agresión en el grupo con modelo fue de 5.20, y la media del grupo sin modelo fue de 3.10. La magnitud del efecto *r* asociada con estos hallazgos fue de 0.688.

Estos breves enunciados informan al lector sobre los patrones generales de los resultados, las medias obtenidas, la significación estadística y la magnitud del efecto. Debe observar la redacción de la frase que incluye el símbolo de la prueba *t*, los grados de libertad y el nivel de significación (probabilidad).

Si los resultados son relativamente directos, se pueden presentar por completo en párrafos escritos. Si el estudio incluyó un diseño complejo, quizás se necesiten figuras y tablas para aclarar la presentación de los resultados.

Tablas y figuras Las tablas generalmente se utilizan para presentar grandes cantidades de datos. Por ejemplo, una tabla puede ser útil para un diseño con varias medidas dependientes; las medias de los diferentes grupos, de todas las medidas dependientes, podrían presentarse en la tabla. Las tablas también son adecuadas para los diseños factoriales. Por ejemplo, en un diseño factorial 2 x 2 x *3,* las 12 medias podrían presentarse en una tabla.

Las figuras se emplean cuando una presentación visual de los resultados ayuda al lector a comprender los hallazgos del estudio. También se pueden uti-

lizar para ilustrar una interacción significativa o para mostrar tendencias a lo largo del tiempo. Para preparar una figura, necesita decidir si presentará la información con una gráfica de pastel, una gráfica de barras o una gráfica lineal. Las gráficas de pastel se utilizan para mostrar porcentajes o proporciones. El pastel completo representa al 100 por ciento y se divide en rebanadas. De esta forma, el todo se divide en grupos o respuestas separadas. Las gráficas de barras son útiles para describir las respuestas de dos o más grupos; por ejemplo, la media de los puntajes de agresión de un grupo con modelo y de un grupo sin modelo, en un experimento. Las gráficas lineales se emplean cuando la variable independiente y la variable dependiente tienen propiedades cuantitativas; por ejemplo, el tiempo de respuesta promedio de dos grupos en los días 1, 2, 3, 4 y 5 de un experimento. Nicol y Pexman (2003) brindan información detallada sobre la creación de figuras y otras representaciones visuales de datos.

En el estilo de la APA, las tablas y las figuras no se presentan en el cuerpo del manuscrito, sino que se colocan al final del documento. Cada tabla y figura aparece en una página separadas. En el texto se hace referencia a una tabla o figura por su número y su descripción del contenido. Nunca haga referencia a la ubicación de una figura, ya que el tipógrafo determina el lugar donde aparece. En la sección de resultados, asevere algo como "la tabla 1 presenta las características demográficas de los encuestados" o "como se muestra en la figura 2, el grupo con modelo. . .". Describa las características importantes de la tabla o figura, en lugar de utilizar comentarios genéricos como "véase la figura 3".

No repita los mismos datos en otro lugar. Una tabla o figura informativa complementa, y no duplica, el texto. El uso de tablas y figuras no disminuye su responsabilidad de establecer con claridad la naturaleza de los resultados en el texto de su reporte.

Cuando se describe un reporte de investigación con propósitos distintos a su publicación; por ejemplo, para cumplir un requisito de un curso o para obtener un grado, es más conveniente colocar cada figura y tabla en una página separada, dentro del cuerpo del documento. Puesto que las reglas sobre la ubicación de las tablas y las figuras pueden variar, verifique el formato apropiado antes de escribir su reporte.

Discusión de los resultados Generalmente *no* es apropiado discutir las implicaciones de los hallazgos en la sección de resultados. Sin embargo, las secciones de resultados y de discusión se pueden combinar si la discusión es breve y si se logra mayor claridad con dicha combinación.

Discusión

La sección de discusión es el lugar adecuado para comentar las implicaciones de los resultados. Una forma de organizar la discusión es iniciar con un resumen del propósito y las expectativas originales del estudio, para luego establecer si los resultados fueron consistentes con sus expectativas. Si los resultados sus-tentan sus ideas originales, debe discutir la forma en que sus hallazgos contri-

buyen al conocimiento del problema investigado. Debe considerar la relación entre sus resultados y las teorías e investigaciones anteriores. Si no obtuvo los resultados esperados, añada posibles explicaciones. Por supuesto, las explicaciones serán muy diferentes, dependiendo de si los resultados obtenidos son opuestos a lo que esperaba o si los resultados no fueron significativos.

Suele ser una buena idea incluir sus propias críticas al estudio. Muchos artículos publicados comentan las limitaciones del estudio. Trate de anticipar lo que los lectores podrían considerar erróneo en su metodología. Por ejemplo, si usted utilizó un diseño de investigación no experimental, podría señalar los problemas de causa y efecto, y la intervención de posibles variables extrañas. En ocasiones, existen errores importantes o insignificantes que podrían corregirse en un estudio subsecuente (si usted tuviera el tiempo, el dinero, etc.). Describa estos errores y sugiera formas de corregirlos. Si existen problemas potenciales para la generalización de los resultados, plantee los problemas y dé las razones por las que piensa que los resultados pueden o no generalizarse.

Es probable que los resultados tengan implicaciones para investigaciones futuras. Si es así, es necesario que discuta la dirección que debe tomar la investigación. También es posible que los resultados tengan implicaciones prácticas, por ejemplo, para la crianza infantil o para mejorar el aprendizaje en el salón de clases. Generalmente, la discusión de estos temas se hace al final de la sección de discusión. Finalmente, probablemente usted deseará incluir un párrafo breve de conclusión que "cierre" el trabajo.

Referencias

La lista de referencias se inicia en una página nueva. Debe contener las citas completas de todas las fuentes mencionadas en el reporte. No omita ninguna fuente de la lista de referencias; asimismo, no incluya fuentes que no hayan sido mencionadas en el trabajo. Más adelante, en este apéndice, se describen los procedimientos exactos para citar las fuentes dentro del cuerpo del reporte y en la lista de referencias. Siga los ejemplos de publicaciones recientes.

Apéndices

No es muy común que se incluyan apéndices en los manuscritos que se envían para ser publicados. El *Manual de Publicaciones* de la APA señala que un apéndice es apropiado cuando un material necesario puede causar distracción en el cuerpo principal del reporte. Ejemplos de apéndices incluyen la muestra de un cuestionario o un instrumento de encuesta, una prueba matemática compleja o una larga lista de palabras utilizadas como estímulos. El apéndice (o varios apéndices) son mucho más apropiados para un proyecto de investigación o para la tesis de un estudiante. Puede incluir la versión completa del cuestionario utilizado, una nueva prueba creada u otros materiales empleados en el estudio. Consulte con su maestro si es apropiado el uso de apéndices en su trabajo. En

caso de incluir alguno, éste se coloca en una página nueva que incluya la palabra "apéndice" en el centro de la parte superior de la hoja.

Nota del autor

La nota acerca del autor comienza con un párrafo que indica la afiliación institucional de los autores. En otro párrafo se pueden incluir los detalles del origen del estudio (por ejemplo, que está basado en la tesis de maestría del primer autor) y los reconocimientos (por ejemplo, financiamientos, colegas que contribuyeron al estudio, etc.). Un párrafo final comienza con "la correspondencia referente a este artículo debe enviarse a...", seguido por la dirección de correo de la persona designada para estos propósitos. Es común que se incluyan las direcciones de correo electrónico en la nota acerca del autor. En los reportes de investigación para cursos universitarios, ésta puede ser innecesaria. La nota acerca del autor se inicia en una nueva página; sin embargo, en ocasiones el maestro (o el editor de una revista) requiere que esta nota se coloque en la parte inferior de la página del título. Esto se hace cuando el documento será sujeto a una *revisión a ciegas*, en la cual el revisor del documento no posee información acerca del autor. En este caso, la página del título se separa del resto del documento antes de la revisión, para enmascarar la identidad del autor.

Notas al pie de página

Si se utilizan notas al pie de página, éstas no se escriben en el cuerpo del texto, sino que se incluyen en una página que se coloca al final del documento. Evite su uso, a menos que sean absolutamente necesarias; suelen distraer a los lectores y la información puede y debe integrarse al cuerpo del documento.

Tablas

Cada tabla debe colocarse en una página separada. Como se señaló antes, el estilo de la APA requiere que las tablas se incluyan al final del reporte, pero es probable que para una clase deba colocarla en páginas separadas dentro del cuerpo del documento. Al preparar una tabla, deje suficiente espacio para que la tabla no aparezca amontonada en una pequeña porción de la página. Defina las áreas de la tabla por medio de líneas horizontales (no utilice líneas verticales). Piense bien el título, de modo que describa de forma precisa y clara el contenido de la tabla. Puede incluir una nota explicativa que muestre los niveles de significación o el rango de los posibles valores de una variable. Antes de crear sus tablas, examine algunas de las que aparecen en un número reciente de las revistas científicas publicadas por la American Psychological Association (Asocia-ción Psicológica Americana).

Figuras

Existen dos reglas especiales del estilo de la APA para la ubicación y preparación de las figuras: 1) las figuras se colocan después de las tablas y 2) antes de las figuras se incluye una página separada con los pies o texto de las figuras. Sin embargo, es probable que ninguna de estas reglas sea necesaria para los reportes o tesis de estudiantes. Tal vez le pidan que coloque cada figura en una hoja separada, en un lugar apropiado del cuerpo del texto, sin necesitar una página *cor*. el texto de las figuras (esto sirve únicamente para la tipografía y para la impresión del reporte). Además, si usted sigue al pie de la letra el estilo de la APA, las páginas con las figuras no incluyen número ni título breve (el número de k figura se anota en el dorso de la hoja con lápiz).

En ocasiones, es necesario preparar las gráficas a mano. Sin embargo, es mucho más fácil utilizar un programa de cómputo para crearlas. La mayoría de los programas de hojas de cálculo, de procesadores de textos y de análisis estadístico incluyen herramientas de graficación. Las variables predictoras e independientes se colocan en el eje horizontal; las variables de criterio y dependientes en el eje vertical. Ambos ejes deben incluir etiquetas. Cuando se imprime la gráfica en una hoja de papel separada, una regla que es el eje horizontal debe tener aproximadamente 12.5 cm de ancho, y el eje vertical alrededor de 9 cm de largo. Si inserta sus gráficas en el texto del reporte (sin aplicar el estilo de la APA), deben ser más pequeñas.

Recuerde que el propósito de una figura es el de aumentar la comprensión de los resultados por medio del uso de una presentación gráfica de los datos. Si la gráfica está atestada de información, confundirá al lector y no servirá a sus propósitos. Planee sus gráficas cuidadosamente para asegurarse de que informa con precisión y claridad al lector. Si está interesado en el tema de la presentación de información en forma gráfica, le recomendamos dos libros escritos por Tufte (1983, 1990). Este autor explora diversas formas para presentar los datos, los factores que contribuyen a la claridad de los datos y las formas en que las gráficas pueden engañar al lector.

Resumen: orden de las páginas

Para resumir, la organización de su reporte es como sigue:

1. Página del título (página 1)
2. Resumen (página 2)
3. Páginas de texto (iniciar en una nueva, página 3)
   - *a)* El título en la parte superior de la primera página inicia la introducción
   - *b)* Método
   - *c)* Resultados
   - *d)* Discusión

4. Referencias (iniciar en nueva página)
5. Apéndice (si se incluye, iniciar en nueva página)
6. Nota del autor (iniciar en nueva página)
7. Notas al pie de página (si se incluyen, iniciar en nueva página)
8. Tablas, con pies o texto de tabla (cada tabla en una página separada)
9. Pies o texto de las figuras (todos los pies juntos en una página separada)
10. Figuras (cada figura en una página separada)

Ahora usted debe tener una idea general de cómo estructurar y redactar su reporte. Lo que resta del apéndice A se enfoca en algunas de las reglas técnicas que pueden ser útiles para la preparación de su trabajo.

EL USO DE ENCABEZADOS

Los documentos que siguen el estilo de la APA utilizan de 1 a 5 niveles de encabezados. Lo más común es emplear el nivel 1 y el nivel 3, aunque a veces también es necesario utilizar el nivel 4. Éstos son como sigue:

(Nivel 1) Encabezado centrado

(Nivel 3) *Encabezado al margen*

El texto inicia con sangría en un nuevo renglón.

(Nivel 4) *Encabezado de párrafo.* El encabezado incluye sangría y el texto inicia en el mismo renglón.

Los niveles 2 y 5 de los encabezados no se describen aquí, debido a que se emplean en documentos más complejos, en los que se presentan experimentos múltiples.

Los encabezados de nivel 1, o centrados, se incluyen para iniciar las secciones principales del reporte: el resumen, el título (en la página 3), el método, los resultados, la discusión, las referencias, etc. Los encabezados de nivel 1 se escriben con mayúsculas iniciales (es decir, la primera letra de cada palabra se escribe con mayúsculas).

Los encabezados de nivel 3, o marginales, se utilizan para dividir las secciones principales en subsecciones. Estos encabezados se colocan en el margen izquierdo, con mayúsculas al inicio de cada palabra. Por ejemplo, la sección del método se divide en por lo menos dos subsecciones: los participantes y el procedimiento. El formato correcto es:

Método

Participantes

La descripción de los participantes se inicia en un nuevo renglón.

Procedimiento

Ahora se describe el procedimiento en forma detallada.

Los encabezados de nivel 4, o de párrafo, se emplean para organizar el material de una subsección. Por ejemplo, la subsección del procedimiento puede separarse en varias categorías para describir las instrucciones dadas a los participantes, la manipulación de la variable independiente, la medición de la variable dependiente y el informe. Cada una de éstas se introduce por medio de un encabezado de párrafo.

Los encabezados de párrafo se inician en un nuevo renglón, con una sangría de 5 espacios o media pulgada. La primera palabra inicia con mayúscula y el resto se escribe con minúscula. El encabezado finaliza con un punto. Toda la información que aparece entre un encabezado de párrafo y el siguiente encabezado (de cualquier nivel) debe estar relacionada con el encabezado de párrafo. Los encabezados de nivel 3 y nivel 4 se escriben en cursivas.

CITAS Y FUENTES DE REFERENCIA

Estilo de citas

Siempre que se refiera a la información reportada por otros investigadores, *debe* identificar las fuentes de manera precisa. Las revistas científicas de la APA utilizan el método de cita autor-fecha. El nombre del autor o autores y el año de publicación se insertan en puntos apropiados. El estilo de cita depende de si los nombres de los autores forman parte del texto o si aparecen entre paréntesis.

Un autor Cuando el nombre del autor está incluido en el texto, incluya la fecha de publicación entre paréntesis, inmediatamente después del nombre:

Markman (1991) encontró que los problemas de pareja pueden conducir a una resolución constructiva del conflicto.

Cuando el nombre del autor no está incluido en el texto, el nombre y la fecha se citan entre paréntesis, al final de una frase introductoria o al final del enunciado:

En un estudio (Markman, 1991), las parejas aprendieron a discutir...

Las parejas muestran tasas más bajas de divorcio y de violencia matrimonial después de la intervención de resolución de problemas (Markman, 1991).

Dos autores Cuando el trabajo tiene dos autores, se incluyen ambos nombres en cada cita. Tanto las citas en el texto, como las que van entre paréntesis, utilizan la conjunción "y" para conectar los nombres de los autores. Cuando los nombres forman parte de un enunciado, y también cuando van dentro de paréntesis, utilice la palabra "y" para unir los nombres de los autores:

Harris y Marmer (1996) reportaron que los padres de familias pobres están menos involucrados con sus hijos adolescentes, que los padres de familias no pobres.

Los padres de familias pobres tienden a pasar menos tiempo con sus hijos adolescentes, que los padres de familias no pobres (Harris y Marmer, 1996).

De tres a cinco autores Cuando un reporte tiene entre tres y *cinco* autores, todos los nombres se citan la primera vez que aparece la referencia. Posteriormente, cite el apellido del primer autor, seguido por la abreviatura *et al.* (y colaboradores), junto con la fecha de publicación. La abreviatura puede utilizarse en las citas que aparecen en el texto o dentro de paréntesis:

Primera cita

Abernathy, Massad y Romano-Dwyer (199 5) reportaron que las adolescentes con baja autoestima son más propensas a fumar, que sus compañeros con alta autoestima.

La investigación sugiere que la autoestima baja es una de las razones por las que las adolescentes se sienten motivadas a fumar (Abernathy, Massad y Romano-Dwyer, 1995).

Citas subsecuentes

Abernathy et *al.* (1995) también examinaron la relación entre el tabaquismo y la autoestima de los adolescentes varones.

En el caso de los varones, no existe una relación entre la autoestima y el tabaquismo, lo que sugiere motivaciones específicas por género del inicio del tabaquismo en la adolescencia (Abernathy et *al.*, 1995).

Otro aspecto referente a las citas subsecuentes es la inclusión de la fecha de publicación cada vez que se nombra a un artículo. En un párrafo, *no* es necesario incluir el año en las citas subsecuentes, siempre y cuando el estudio no se confunda con otros estudios citados en el reporte.

Citas en un párrafo

En un estudio reciente del tiempo de reacción, Yokoi y Jones (1998)...

Yokoi y Jones también reportaron que...

La fecha de publicación debe incluirse cuando las citas subsecuentes se encuentren en otro párrafo o en otra sección del reporte.

Seis o más autores Ocasionalmente, se cita una referencia de seis o más autores. En este caso, utilice la abreviatura *et al.* después del apellido del primer autor, en *cada* cita. Aun cuando no se listan todos los nombres de los autores en el texto, la cita en la sección de referencias del reporte debe incluir los nombres de los primeros seis autores, seguido por la abreviatura *et al.*

Referencias sin autor Cuando un artículo no tiene autor (por ejemplo, algunos artículos de periódicos o revistas), se citan las primeras dos o tres palabras del título entre comillas, seguidas por la fecha de publicación:

Cita en la lista de referencias

El tabaquismo de los padres mata 6200 niños anualmente, afirma un estudio. (1997, julio 15). *Orange County Register,* p. 11.

Cita en el texto

```
    En un articulo sobre el tabaquismo
("Tabaquismo paterno", 1997), datos recuperados
de...
```

Múltiples trabajos en el mismo paréntesis Una forma conveniente para citar varios estudios sobre el mismo tema, o varios estudios con hallazgos similares, es referirlos como una serie dentro del mismo paréntesis. Cuando dos o más trabajos son del mismo autor(es), debe reportarlos de acuerdo al orden del año de publicación, utilizando comas para separar las citas:

```
Sehaie y Willis (1986, 1993) encontraron...

Investigaciones anteriores (Sehaie y Willis, 1986,
1993) indican...
```

Cuando cite dos o más trabajos de distintos autores en el mismo paréntesis, ordénelos alfabéticamente y separe las citas con punto y coma:

```
Las investigaciones con familias que padecen
crisis económicas, reportan de forma consistente
que las niñas reaccionan con problemas de
internalización, mientras que los niños responden
con problemas de externalización (Con-ger, Ge,
Eider, Lorenz y Simons, 1994; Flanagan y
Eccles, 1993; Lempers, Clark-Lempers y Simons,
1989).
```

Estilo de la lista de referencias

El *Manual de Publicaciones* de la APA ofrece ejemplos de 95 distintos formatos de referencias para los artículos de revistas científicas, libros, capítulos de libros, reportes técnicos, presentaciones en congresos, tesis, páginas web y videos, entre otros. Aquí se incluyen sólo algunos de ellos. Cuando tenga alguna duda sobre la elaboración de una referencia, consulte el manual de la APA. El formato general para una lista de referencias es el siguiente:

1. Las referencias se citan en el orden alfabético del apellido del primer autor. No categorice las referencias de acuerdo a su tipo (es decir, libros, artículos de revistas científicas, etc.). Observe los espacios en los nombres de los autores en los ejemplos.
2. Los elementos de una referencia (nombres de los autores, título del artículo, fecha de publicación) se separan con puntos.

El primer renglón de cada referencia se coloca en el margen izquierdo y los renglones siguientes incluyen una sangría denominada "sangría colgante". La referencia aparece de la siguiente manera:

Davis, J. L., y Rusbult, C. E. (2001). Attitude alignment in close relationships. *Journal of Personality and Social Psychology, 81,* 65-84.

Cada referencia se anota en un nuevo renglón (piense en cada referencia como un nuevo párrafo). La mayoría de los procesadores de textos permiten dar formato con facilidad al párrafo con sangría colgante, de modo que usted no tiene que insertar los espacios manualmente en el segundo renglón y subsecuentes: si utiliza Microsoft Word, inicie el párrafo con ctrl-t (presione las teclas control y t al mismo tiempo); con WordPerfect, inicie el párrafo de la referencia con ctrl-F7.

Formato para artículos de revistas científicas La mayoría de las revistas están organizadas de acuerdo con el volumen y al año de publicación (por ejemplo, el volumen 57 de la revista *American Psychologist* consiste de los números de la revista publicados en 2002). Una confusión común es la inclusión del número del ejemplar, además del número del volumen. La regla es sencilla: si los ejemplares en un volumen tienen paginación consecutiva a lo largo del volumen, *no* se debe incluir el número del ejemplar. Si cada ejemplar de un volumen inicia en la página 1, debe incluirse el número del ejemplar. A continuación se muestran ejemplos específicos.

En la lista de referencias, el nombre de la revista y el número del volumen se anotan en cursivas. Además, únicamente la primera letra de la primera palabra del título del artículo se escribe con mayúscula (con excepción de los nombres propios). Tome en cuenta que en las referencias en inglés, después de los dos puntos se inicia también con mayúsculas; en español, después de los dos puntos se inicia con minúsculas. He aquí algunos ejemplos:

Una autor-sin número de ejemplar

Newby, T. J. (1991). Classroom motivation strategies: Strategies of first-year teachers. *Journal of Educa-tional Psychology, 83,* 195-200.

Dos autores-uso del número de ejemplar

Greenwald-Robbins, J., y Greenwald, R. (1994). Environmental attitudes conceptualized through developmental theory: A qualitative analysis. *Journal of Social Issues,* 50(3), 29-47.

Formato para libros Cuando se cita un libro, el título se escribe en cursivas. Únicamente la primera palabra del título se escribe con mayúscula; sin embargo, los nombres propios también se escriben con mayúsculas y se aplica la regla de los dos puntos ya mencionada. Después del título se anota la ciudad de la publicación y la empresa que lo publica. Si no se *conoce* la ciudad, se incluye la abreviatura de dos letras del servicio postal del estado de Estados Unidos (por ejemplo, AZ, NY, MN, TX).

Libro de un autor

Uba, L. (1994). *Asían Americans: Personality patterns, identity, and mental health.* New York: The Guilford Press.

Libro de una autor-segunda o posterior edición

McAdoo, H. P (1988). *Black famllles* (2a ed.). Newbury Park, Ca: Sage.

Libro editado

Huston, A. H. (Ed.). (1991). *Children In poverty: child development and public policy.* New York: Cambridge University Press.

Formato para artículos en libros editados Para libros editados, la referencia inicia con los nombres de los autores del artículo, y no del libro. Después se anota el título del artículo y luego el nombre del editor o editores del libro, el título del libro, los números de las páginas del artículo y la fecha de publicación de libro. Sólo el título del libro se escribe en cursivas, y la primera letra del artículo y del libro se escribe con mayúscula. He aquí algunos ejemplos.

Un editor

Brown, A. L., y Campione, J. C. (1994) Guided discovery in a community of learners. En K. McGilly (Ed.), *Classroom lessons: Integratlng cognltive theory and classroom practice* (pp. 229-270). Cambridge MA: MIT press.

Dos editores

Bates, J., Bayles, K., Bennet, D., Ridge, B., y Brown, M. (1991). Origins of externalizing behavior problems at eight years of age. En D. Pepler y K. Rubin (Eds.), *The*

development and treatment of childhood aggression (pp. 93-120). Hillsdale, NJ: Erlbaum.

Capítulo de un libro de una serie de volúmenes

Kagan, J. (1992). Temperamental contributions to emotion and social behavior. En M. S. Clark (Ed.), *Review of personality and social psychology: vol. 14. Emotions and social behavior* (pp. 99-118). Newbury Park, CA: Sage.

Formato para "artículos populares" Los estilos de referencias que se muestran a continuación, se utilizan para artículos de revistas *y* periódicos. Como regla general, los artículos populares de la prensa se emplean moderadamente (por ejemplo, cuando no es posible encontrar artículos científicos sobre un tema o para dar un ejemplo de un evento que está relacionado con el tema).

Revista-paginas continúas

Begley, S. (1995, 27 de marzo). Gray matters. *Newsweek, 125,* 48-54.

Periódico-sin autor

10-year-old is youngest to gradúate from college. (1994, 6 de junio). *Orange County Register,* p. 15.

Periódico-paginas discontinuas

Colé, K. C. (1995, I° de mayo). Way the brain works may play a role in bias, experts say. *Los Angeles Times,* pp. Al, A18.

Formato para trabajos y sesiones de carteles presentadas en conferencias Ocasionalmente, necesitará citar un trabajo no publicado o una sesión de carteles que fueron presentada en un congreso profesional. A continuación se muestran dos ejemplos:

Trabajo

Kee, D. W., McBride, D., Neale, P., y Segal, N. (1995, noviembre). *Manual and cerebral laterality in hand-discordant monozygotic twins.* Trabajo presentado en

el congreso anual de la Psychonomic Society, Los Angeles, CA.

Sesión de carteles

Roach, M. A., Barratt, N. S., y Miller, J. F. (1997, abril). *Maternal adaptation over time to children with Down Syndrome and typically developing children.* Sesión de carteles presentada en el congreso anual de la Society for Research in Child Development, Washington, DC.

Fuentes secundarias A veces es necesario citar un artículo, libro o capítulo de un libro que se leyó a través de un libro de texto, un resumen o la revisión del libro. Aunque siempre es preferible leer las fuentes primarias, en ocasiones se debe citar una fuente secundaria.

Suponga que desea citar un artículo del que se habla en un libro. Cuando se refiera al artículo en su trabajo, debe indicar que estaba citado en el libro. En el siguiente ejemplo, el trabajo de Conway y Pleydell-Pearce es la fuente primaria:

Conway y Pleydell-Pearce (citado en Woll, 2002) sugirieron que la memoria autobiográfica...

En la lista de referencias, al final del trabajo, simplemente anota la referencia de la fuente secundaria que utilizó (en este caso, la cita de Woll):

Woll, S. B. (2002). *Everyday thinking: memory, reasoning, and judgment in the real world.* Mahwah, NJ: Erlbaum.

Si hace referencia a un resumen que obtuvo de la base de datos *PsycINFO*, sencillamente cite al autor o autores en su trabajo, como lo haría normalmente. Sin embargo, la referencia necesita indicar que la cita está basada en un resumen. Siga este ejemplo en el que se da la información de la referencia, pero también se indica que el resumen se recuperó en la fecha que se especifica:

Bieman-Copland, S., y Ryan, E. B. (2001). Social perceptions of failures in memory monitoring. *Psychology and Aging, 16,* 3 5 7-361. Resumen recuperado el 20 de septiembre del 2001, de la base de datos *PsycINFO.*

Formato de referencias de fuentes electrónicas

Durante los últimos años, la cantidad de información disponible en Internet se ha incrementado de forma sorprendente. La American Psychological Association (Asociación Psicológica Americana) ha creado lineamientos para citar información obtenida por este medio. Estos lineamientos están publicados en el manual de la APA, y también en http://www.apastyle.org. Los formatos más actuales se incluyen aquí, pero es muy probable que cambien conforme la tecnología continúe desarrollándose.

Cita de una página web

En ocasiones, usted querrá citar una página web en su trabajo, sin referirse a un documento específico. En este caso, únicamente proporcione la dirección de la página; no es necesario incluirlo en su lista de referencias. Por ejemplo, lo siguiente podría aparecer en el texto de su documento:

El libro *Methods In Behavioral Research* tiene una útil página web para los estudiantes (http://methods.fullerton.edu).

Cita de documentos web específicos/páginas de internet

Los métodos para citar documentos web específicos son iguales a las citas de cualquier documento. Siga las reglas que ya fueron descritas y anote el autor y el año de publicación. Por ejemplo, para citar el documento web que contiene estos lineamientos, el texto diría:

La American Psychological Association (2002) ha preparado un conjunto de lineamientos para citar fuentes electrónicas.

En la lista de referencias, la cita real del documento web sería:

American Psychological Association. (2001). *Electronic references.* Recuperado el I° de noviembre, 2002, de http://www.apastyle.org/elecsource.html

Observe que la cita trata al documento web como una publicación sin fecha y sin editor. Lo importante es proporcionar la información de la fecha en que se obtuvo el documento, el lugar exacto y el nombre del archivo (URL) del documento. La regla general es proporcionar información suficiente de la cita para permitir que un lector localice la información. En ocasiones, es difícil anotar

toda la información que puede obtener en una publicación impresa regular; por ejemplo, es probable que el documento no tenga fecha; para esto puede utilizar la abreviatura s.f. para indicar que se trata de información sin fecha.

Existe una regla importante acerca del URL (ubicación) del documento que está citando. Se acepta anotar la ubicación en dos renglones, si no cabe en uno solo. Sin embargo, nunca se debe insertar un guión, ya que éste no forma parte de la dirección y dificultaría a lector la obtención de la información.

Las revistas científicas y periódicos electrónicos se están volviendo más comunes. A continuación presentamos una cita de una revista científica electrónica de psicología llamada *Prevention & Treatment.*

Beutler, L. E. (1° de septiembre, 2000). Empirically based decisión making in clinical practice. *Prevention & Treatment, 3,* artículo 27. Recuperado el 1° de marzo de 2001, de http://journals.apa.org/prevention/volume3/_pre0030027a.html

Observe que la revista ofrece información sobre el número de volumen, pero no existe número de página, debido a que el documento es un archivo sin páginas separadas. En este caso no se requiere de un punto al final de la cita.

ABREVIATURAS

Los trabajos que utilizan el estilo de la APA no emplean demasiadas abreviaturas ya que pueden ser distractoras, debido a que el lector debe intentar constantemente traducir la abreviatura a su significado. Sin embargo, el estilo de la APA permite el uso de las abreviaturas que se aceptan como términos en el diccionario. Éstas incluyen CI, MOR, PES y SIDA.

Ciertos términos bien conocidos pueden abreviarse cuando facilitan la lectura, pero su significado debe aclararse cuando se utiliza por primera vez en el documento. Algunos ejemplos de abreviaturas comunes son

MMPI	Inventario multifásico de la personalidad de Minnesota
MCP	Memoria de corto plazo
EC	Estímulo condicionado
TR	Tiempo de reacción
CVC	Consonante-vocal-consonante
ANOVA	Análisis de varianza

En ocasiones, los términos estadísticos se utilizan de manera abreviada o simbólica. Estos siempre se escriben en cursiva. Por ejemplo,

M	media
DE	desviación estándar
Md	mediana
gl	grados de libertad
n	número de individuos en un grupo o condición experimental
N	número total de participantes
p	nivel de probabilidad (significación)
SC	suma de cuadrados
CM	cuadrado medio
F	valor de *F* en el análisis de varianza
r	coeficiente de correlación de Pearson
R	coeficiente de correlación múltiple

Algunas de las abreviaturas científicas de diversas unidades de medición, que se utiliza con frecuencia, son:

cm	centímetro
g	gramo
hr	hora
pulg	pulgada
kg	kilogramo
km	kilómetro
m	metro
mg	miligramo
min	minuto
ml	mililitro
mm	milímetro
ms	milisegundo
s	segundo

Finalmente, ciertas abreviaturas del latín y del inglés de la edad media, se utilizan regularmente en los trabajos. Algunas de estas abreviaturas y su significado son:

cf.	comparar	(del latín *confer)*
e-g-,	por ejemplo	(del latín *exempli gratia)*
etc.	etcétera	(del latín *et cetera)*
i.e.,	es decir	(del latín *id est)*

viz.	a saber
vs.	Versus

ALGUNAS CONSIDERACIONES GRAMATICALES

Palabras y frases de transición

Una forma de crear un reporte de investigación con escritura clara, es poner atención a la manera en que se conectan los enunciados dentro de un párrafo, y la conexión de los párrafos dentro de una sección. Las transiciones entre los enunciados y párrafos deben ser suaves y consistentes con la línea de razonamiento. En esta sección se describen algunas palabras y frases de transición de uso común, y sus funciones.

Adverbios Los adverbios pueden utilizarse como palabras introductorias en los enunciados. Sin embargo, debe emplearlos con su significado correcto.

Adverbio	***Significado implícito***
(Des)afortunadamente	Es una fortuna que...
Similarmente	De forma similar...
Ciertamente	Es cierto que...
Claramente	Está claro que...

Palabras que sugieren contraste Algunas palabras y frases sugieren un contraste o contradicción entre lo que está escrito antes y lo que está escrito ahora:

Entre enunciados	***Dentro de los enunciados***
En contraste,	mientras que
Por otro lado,	aunque
Sin embargo,	pero

Las palabras de la lista izquierda se refieren al enunciado anterior. Las palabras de la lista derecha conectan frases dentro de un enunciado; es decir, se refieren a otro punto en el mismo enunciado.

Palabras que sugieren una serie de ideas Las palabras y frases que sugieren que la información que se da después de la palabra de transición está relacionada con o es similar a la información en el enunciado, son:

Primero	Además	Por último	Es más
Segundo	Adicionalmente	Finalmente	Lo que es más
Tercero	Entonces	Asimismo	Otro

Palabras que sugieren implicaciones Estas palabras y frases indican que la información que sigue a la palabra de transición está implicada por, o proviene de la información anterior:

Por lo tanto	Si...entonces
Se sigue que	Así
En conclusión	Entonces

Cuando utilice palabras de transición, asegúrese de que implican el significado que usted quiere darles. Su uso indiscriminado en el inicio de los enunciados provoca confusión en el lector y distorsiona sus propósitos.

Palabras y frases problemáticas

"Que" versus "cual" *Que y cual* son pronombres relativos que introducen cláusulas subordinadas y reflejan la relación de la cláusulas subordinadas con la cláusula principal. Las cláusulas que incluyen el término *que* se conocen como restrictivas y son esenciales para precisar el significado del enunciado; las cláusulas con el término *cual* no son restrictivas y solamente añaden información. Observe los distintos significados del mismo enunciado al utilizar *que* y *cual.*

> Los ratones que tuvieron un buen desempeño en el primer ensayo se utilizaron en el segundo ensayo.
>
> Los ratones, los cuales tuvieron un buen desempeño en el primer ensayo, se utilizaron en el segundo ensayo.

El primer enunciado establece que únicamente los ratones que tuvieron un buen desempeño en el primer ensayo, se utilizaron en el segundo. El segundo enunciado afirma que todos los ratones se utilizaron en el segundo ensayo, y que también todos se desempeñaron bien en el primer ensayo.

"Mientras" versus "desde" *Mientras* y *desde,* son conjunciones subordinadas que también introducen cláusulas subordinadas. Para aumentar la claridad en la escritura científica, el manual de la APA sugiere que *mientras* y *desde,* se utilicen únicamente para referirse al tiempo. *Mientras,* se emplea para describir eventos simultáneos y *desde,* para referirse a eventos subsecuentes:

> Los participantes esperaron juntos mientras calificaban sus pruebas de personalidad.
>
> Desde la aparición del estudio de Eider (1974), se han publicado muchos estudios sobre este tema.

El manual de la APA sugiere el uso de otras conjunciones para vincular frases que no describen eventos temporales. *Aunque, aun cuando, mientras que y pero* pueden emplearse en lugar de *mientras, y porque* debe sustituirse por *desde.*

Incorrecto: Mientras que el estudio estaba bien diseñado, el reporte estaba mal escrito.

Correcto: Aunque el estudio estaba bien diseñado, el reporte estaba mal escrito.

REPORTE DE CIFRAS Y ESTADÍSTICOS

Prácticamente todos los documentos de investigaciones reportan cifras: número de participantes, número de grupos, valores de estadísticos como *t, F o r.* ¿Se deben utilizar números (por ejemplo, *"43")* o se deben emplear palabras (por ejemplo, *"cuarenta y tres")?* La regla general es utilizar palabras cuando se expresan los números del cero al nueve, y números del 10 en adelante. Sin embargo, existen algunas aclaraciones importantes.

Si usted inicia un enunciado con un número, debe emplear palabras, incluso si el número es 10 o mayor (por ejemplo, *"Ochenta y cinco estudiantes de maestría participaron en el estudio").* Es raro que se inicie un enunciado con un número, especialmente si se trata de números grandes. Por lo tanto, debe tratar de revisar el enunciado para evitar el problema (por ejemplo, *"Los participantes fueron 85 estudiantes inscritos en clases para convertirse en maestros").*

Cuando se comparan cifras mayores y menores que 10 en el mismo enunciado, se utilizan números en ambos (por ejemplo, *"los participantes leyeron 8 o 16 párrafos").* Sin embargo, el siguiente enunciado contiene una *mezcla* apropiada de números y palabras: *"los participantes leyeron ocho párrafos y después respondieron 20 preguntas de opción múltiple".* El enunciado es correcto debido a que los párrafos y las preguntas son entidades distintas y no están siendo comparadas.

Cuando se reporta un porcentaje, utilice números seguidos por el signo de porcentaje, excepto cuando inicie un enunciado o se reporten las cantidades de manera esporádica dentro de los párrafos, para estas dos excepciones se puede escribir la palabra "por ciento". Esto se hace sin importar si el número es menor que 10 (por ejemplo, *"sólo el 6% de los juegos de computadora atrajo a las mujeres")* o mayor que 10 (por ejemplo, *"con el uso de esta técnica, el 85% de los participantes mejoró su desempeño").*

Para describir edades siempre se utilizan números (por ejemplo, *"los niños de 5 años de edad"),* así como también para describir los puntos en una escala (por ejemplo, *"un 3 en una escala de 5 puntos"),* unidades de medición (por ejemplo, *"los niños se pararon a 2 metros del blanco"),* tamaños muéstrales (por ejemplo, *"6 niñas y 6 niños fueron asignados a cada condición del estudio") y* estadísticos (por ejemplo,

"la puntuación media del grupo sin modelo fue de 3.10"). Una rara, pero sensible excepción a esta regla, ocurre cuando dos tipos de números diferentes deben aparecer juntos. Un ejemplo es *"los maestros identificaron a quince niños de 7 años, como los más agresivos".* Este enunciado evita la yuxtaposición de dos números.

Para el signo de multiplicación, utilice la x minúscula o el símbolo de multiplicación empleado por su procesador de textos. Esto se aplica cuando se describen una operación matemática o un diseño factorial (por ejemplo, un diseño 2 X 2). Para él signo menos, utilice un guión con un espacio antes y después de éste.

Finalmente, es necesario que conozcan la forma en que se presentan los resultados estadísticos en un trabajo. Como se señaló anteriormente, los términos estadísticos se abrevian y se escriben en cursivas (por ejemplo, *M, r, t, F).* Además, cuando se reporten los resultados de una prueba de significación estadística, anote el nombre de la prueba, los grados de libertad, el valor del estadístico de prueba y el nivel de probabilidad. A continuación se presentan dos ejemplos de enunciados que describen resultados estadísticos:

> Tal como se predijo, los participantes en la condición de alta ansiedad tardaron más tiempo en reconocer las palabras ($M = 2.63$) que lo que tardaron los individuos en la condición de baja ansiedad ($M = 1.42$), $t(20) = 2.54$, $p = 0.019$.

> Las puntuaciones de satisfacción en el empleo se correlacionaron de forma significativa con la satisfacción marital, $r(50) = 0.48$, $p < 0.001$.

Recuerde que las probabilidades exactas se reportan tal como aparecen en los resultados del análisis estadístico que le da la computadora. Sin embargo, en ocasiones la computadora no reporta probabilidades muy bajas, menores de 0.001; en estos casos, se acepta el uso del símbolo < (menor que).

Si su impresora no puede producir un símbolo específico, puede dibujarlo con tinta negra. Ponga atención a la forma en que se describen los estadísticos en los artículos que lea. Encontrará que puede variar las descripciones de sus resultados para que se ajusten mejor a sus datos y a su presentación; también varíe su construcción de enunciados.

CONCLUSIÓN

Cuando haya completado su reporte de investigación, deberá sentirse orgulloso de su esfuerzo. Usted habrá tomado en cuenta las investigaciones anteriores sobre un problema, realizado un proyecto de investigación, analizado los resultados y reportado los hallazgos. Un esfuerzo de investigación como éste puede

dar como resultado una publicación o una presentación en un congreso. Sin embargo, ésta no es la parte más importante de su investigación. Lo más importante es que ha adquirido nuevos conocimientos y que se ha despertado su curiosidad, de modo que deseará aprender más.

EJEMPLO DE UN REPORTE

El resto de este apéndice *consiste* de un reporte que fue publicado en una revista científica. Tiene la intención de ser una guía útil cuando usted escriba sus propios reportes siguiendo el estilo de la APA. Las notas al margen señalan elementos importantes del estilo de la APA. Lea cuidadosamente el manuscrito, ponga especial atención al formato general, y asegúrese de que comprende las reglas respecto a la numeración de las páginas, los encabezados de las secciones, las citas de referencias y el formato de las figuras. Escribir el primer reporte de investigación siempre constituye un reto. Esto irá siendo más fácil conforme lea las investigaciones de otros autores y practique escribiendo sus propios reportes.

Jennifer L. Knight y Traci A. Giuliano amablemente dieron su permiso para reproducir su artículo, para ejemplificar algunos elementos del estilo de la APA. Los comentarios al margen lo alertan de características del estilo de la APA que necesita conocer cuando escriba sus propios trabajos. Sin embargo, esté consciente de que cada trabajo incluye tipos ligeramente diferente de información, dependiendo del tema en particular, el método y los resultados. Su trabajo podrá seguir los lineamientos generales del estilo de la APA, pero muchos de los detalles están determinados por las necesidades de su estudio.

Notará que el título de este trabajo sobrepasa las 10 a 12 palabras recomendadas por la APA. Un título más breve (pero menos intrigante) podría ser *Consecuencias de las representaciones de estereotipos por género, de atletas hombres y mujeres, en los medios impresos.*

Cada página tiene un encabezado con la identificación del trabajo y el número de la página. La identificación del trabajo se forma con las primeras palabras del título.

El es un Laker 1

El título principal es un título abreviado que aparece en la parte superior de cada página, cuando se publica el artículo. Incluye un máximo de 50 espacios y se escribe con mayúsculas.

Titulo principal: REPRESENTACIONES POR GÉNERO DE ATLETAS EN LOS MEDIOS

El título, el autor o autores y su afiliación aparecen centrados en la mitad superior de la página.

El título suele incluir de 10 a 12 palabras.

Él es un Laker; ella es una "observadora": Las consecuencias de las representaciones de estereotipos por género, de hombres y mujeres atletas, en los medios impresos

Jennifer L. Knight y Traci A. Giuliano
Southwestern University

Todo el documento se escribe a doble espacio.

Los márgenes son de al menos 2.5 cm (generalmente son apropiados los márgenes establecidos de forma automática por el procesador de textos).

Él es un Laker 2

Resumen

Aun cuando la investigación cualitativa muestra que la cobertura de los deportes femeninos suele enfocarse en el atractivo de las atletas (sin tomar en cuenta su fortaleza atlética), escasean las investigaciones cuantitativas que examinen los efectos de esta cobertura sobre la forma en que la gente percibe a los atletas. Como parte de un diseño 2x2 entre sujetos, 92 estudiantes de licenciatura (40 hombres, 52 mujeres) leyeron un artículo que hablaba acerca del perfil ficticio de un atleta olímpico, con un enfoque en el atractivo del atleta o en sus habilidades atléticas. De forma interesante, a los participantes no les gustaron ni tuvieron una impresión favorable de los artículos que hablaban del atractivo de hombres y mujeres atletas, cuando dicho atractivo fue el principal enfoque del artículo. Estos hallazgos sugieren que los medios deben estar conscientes de las impresiones que pueden resultar de representar a los atletas en términos de sus atributos personales y no de sus logros atléticos.

El resumen inicia en una nueva página

La palabra "Resumen " va en el centro y la primera letra se escribe con mayúsculas.

No se emplean sangrías en el resumen.

Por lo general, el resumen no excede de 120 palabras.

Utilice un solo espacio después de un punto, una coma, dos puntos y punto y coma.

El cuerpo principal del documento inicia en la página 3. El cuerpo del documento incluye las secciones de introducción, método, resultados y discusión.

Centre el título y después empiece a escribir la introducción. No incluya los nombres de los autores.

Utilice el apellido del autor y la fecha para las citas.

Él es un Laker; ella es una "observadora": las consecuencias de las representaciones de estereotipos por género, de hombres y mujeres atletas, en los medios impresos En una era en la que los deportes profesionales masculinos se caracterizan por contratos millonarios en dólares, por el cierre de sindicatos de jugadores, por el uso ilegal de esteroides y por una mentalidad individualista, los disgustados fanáticos de los deportes están orientándose a los deportes profesionales femeninos para entretenerse. De hecho, ligas como la Women's National Basketball Association (WNBA), la Ladies' Professional Golf Association (LPGA), la Women's Pro Softball League (WPSL), el Women's Pro Tennis Tour (WTA) y la Women's United Soccer Association (WUSA) constituyen una señal de bienvenida para los fanáticos que buscan deportes de equipo, asientos poco costosos y estrellas deportivas accesibles (Wulf, 1997). Además del floreciente campo de los deportes profesionales femeninos, los Juegos Olímpicos también son escaparate de mujeres atletas exitosas. En los juegos de Atlanta de 1996, los equipos femeninos estadounidenses ganaron medallas de oro en gimnasia, fútbol soccer, softbol y basquetbol (los equipos de softbol y basquetbol repitieron sus títulos en los juegos de Sydney 2000) . Sus contrapartes invernales también tuvieron éxito en los juegos de Nagano en 1998, donde el equipo femenino de hockey ganó el primer lugar, y con el

triunfo de las estrellas individuales Picabo Street, Tara Lipinski y Christine Witty.

Las atletas femeninas que compiten a nivel interescolar e intercolegial, también han avanzado a grandes pasos. La decisión de la corte con titulo IX, de 1972, obliga a todos los programas que reciben fondos federales, incluyendo atletismo, a proporcionar un trato igualitario y la misma oportunidad de participación a hombres y mujeres. Para los programas deportivos, esto implicó que las escuelas preparatorias y las universidades públicas asignaran cantidades equivalentes de tiempo y dinero para becas de hombres y mujeres atletas, para su reclutamiento, instalaciones, material, viajes y servicios (Curtis y Grant, 2001). En parte debido a estas mejores oportunidades, la participación de niñas y mujeres en los deportes ha alcanzado su nivel más alto de todos los tiempos. Mientras que en 1971 sólo 1 de cada 27 niñas participaba en atletismo en la escuela secundaria y preparatoria, en 1997 participaron más de 1 de cada 3 (Women's Sports Foundation, 1998).

Aun cuando la participación de las mujeres en deportes profesionales, olímpicos, intercolegiales e interescolares ha alcanzado niveles sin precedentes, las investigaciones muestran que la cobertura que hacen los medios de las mujeres atletas todavía es mucho menor que la de los hombres (por ejemplo, Duncan, Messner y Williams, 1991; Fink, 1998; Tugle y Owen, 1999). Por

Anote los nombres de todos los autores en la primera cita. Cuando haya de tres a cinco autores utilice "et al.*" en las citas subsecuentes. Vea el texto para las excepciones con más de cinco autores.*

Cuando cite varias referencias al mismo tiempo, sepárelas con punto y coma.

Utilice letras cursivas para los títulos de libros, revistas científicas y revistas populares.

Observe que no se utiliza un apóstrofo en las décadas, como en 195Os.

Observe las reglas para reportar cifras: 4 y 53 se utiliza en el mismo enunciado, por lo que el 4 no se escribe con letra.

ejemplo, en la revista *Sports Illustrated*, sólo 4 de 53 números de 1996 incluyen mujeres en la portada (Women's Sports Foundation, 1998). Un estudio longitudinal de los artículos de esta misma revista, desde la mitad de la década de 1950 hasta finales de la década de 1980, también reveló que la popular revista incluye un número mucho menor de columnas y fotografías por artículo de deportes femeniles, en comparación con los de hombres (Lumpkin y Williams, 1991; Salwen y Wood, 1994) . La cobertura televisiva de los juegos olímpicos, tanto en 1992 (Higgs y Weiller, 1994) como en 1996, propuesto el "año de los deportes femeninos" (Eastman y Billings, 1999), exhibió un patrón similar. Incluso la cobertura de los deportes colegiales y de educación media presenta un sesgo por género; los niños son el centro de más y más largos artículos, que las niñas (por ejemplo, Sagas, Cunningham, Wigley y Ashley, 2000; Woodcock, 1995). En efecto, esta "aniquilación simbólica" (Gerbner, 1972) de los deportes femeninos, por parte de los medios, comunica la idea falsa de que los deportes femeninos son inferiores y no tan notables como los deportes masculinos.

La cobertura de los deportes femeninos es inferior que la de los hombres no sólo en cantidad, sino también en calidad (e.g., Birrell y Theberge, 1994; Daddario, 1998; Duncan, 1990; Duncan y Messner, 2000; Kane y Parks, 1992). Los comentaristas y escritores de deportes con frecuencia aluden o se refieren de forma explícita

e..g.,) significa "por ejemplo" i. e., significa "es decir".

al atractivo de mujeres atletas, asi como también a sus emociones, feminidad y heterosexualidad (lo que efectivamente indica a la audiencia que el estereotipo del papel del género es más notorio que su papel atlético), mientras que los hombres atletas son descritos como poderosos, independientes, dominantes y valiosos (Hilliard, 1984; Messner, 1988; Sabo y Jansen, 1992). Debido a que la participación competitiva en los deportes es inconsistente con el papel femenino prescrito por la sociedad, la cobertura que los medios hacen de las mujeres atletas aparenta intentar protegerlas del rechazo (o, desde un punto de vista más cinico, le dan al público lo que creen que "desea"), cuando se hace énfasis en otros aspectos de sus "aspectos de mujer", como su atractivo (Kane, 1996). Por ejemplo, aun cuando Gabrielle Reese, Anna Kournikova, Katarina Witt y Jan Stephenson son atletas excepcionales, con frecuencia los medios se enfocan en su atractivo fisico, problema que es mucho menos común en los hombres atletas. En efecto, los medios tienden a representar a las atletas primero como mujeres (es decir, cuando se enfocan en su cabello, sus uñas, su ropa y su atractivo fisico) y después como atletas; sin embargo, los hombres atletas suelen ser representados únicamente en términos de sus habilidades atléticas (Boutilier y SanGiovanni, 1983).

Esta imagen trivial de las mujeres atletas es consistente con la teoria de los esquemas, que propone que la gente posee estructuras

cognoscitivas implícitas que les proporcionan expectativas cuando procesan información (Fiske y Taylor, 1991). Uno de los estereotipos más dicotómicos y con mayor influencia social es el del género (Burn, O'Neal y Nederend, 1996). La teoría del esquema de género argumenta que las personas socializan (por ejemplo, por medio de los padres, maestros, compañeros, juguetes y los medios populares de comunicación) y creen que las diferencias de género son significativas y que vale la pena mantenerlas (Bem, 1981). A pesar de que actualmente existe una mayor variabilidad dentro que entre los sexos, persiste el concepto de las diferencias distintivas y exclusivas entre los géneros (Martin, 1987).

Cuando la gente viola nuestros bien arraigados esquemas (como una mujer que maneja un camión o un hombre que hace labores secretariales), es percibida de forma más negativa, que las personas que son consistentes con los esquemas (por ejemplo, Knight, Giuliano y Sanchez-Ross, 2001; Rosenthal y Jacobson, 1968). Puede ser, entonces, que los hombres ya estén representados en los medios como atletas, primero porque ser un atleta es consistente con el papel masculino tradicional (Coakley y White, 1992). Sin embargo, si una mujer es atleta, esto contradice el papel femenino convencional y, por lo tanto, la cobertura de los medios enfatiza otros aspectos de sus "características de mujer" (como su atractivo físico). Como consecuencia, las coberturas sobre los hombres atletas son libres de enfocarse en sus

logros atléticos, mientras que las representaciones de mujeres atletas se enfocan en aspectos de su feminidad, quizás para hacer que ellas aparezcan de una forma más consistente con su papel de género.

Está bien establecida la imagen trivial de los deportes femeninos en los medios, pero los investigadores aún tienen que indagar de manera empírica la forma en que las diferentes representaciones de los hombres y mujeres atletas afectan la percepción que el público tiene de ellos. A pesar de que los investigadores han especulado acerca de la forma en que las creencias de las personas pueden verse influidas por coberturas prejuiciadas (por ejemplo, Fasting, 1999), existen muy pocos estudios sobre las consecuencias reales de estas representaciones distintas, asi como sobre la magnitud en que los medios pueden realmente influir en la percepción que la gente tiene de los atletas. Además, miembros de los medios de comunicación argumentan que ellos solamente dan la cobertura que "el público desea", aunque esto también debe apoyarse en investigaciones empíricas. *En* otras palabras, aún se desconoce hasta qué punto los medios reflejan única o activamente la opinión pública. El propósito de la presente investigación fue el de abordar estas preguntas que estaban sin respuesta en la literatura deportiva. Para explorar cómo es que las representaciones consistentes e inconsistentes con el género de los atletas, afectan las percepciones de la gente, se

diseñó un perfil olímpico hipotético, en el que el punto central del articulo consistió en el atractivo fisico de hombres o mujeres atletas (la tipica representación femenina), o las habilidades atléticas (la tipica representación masculina).

En general, se predijo que las mujeres atletas, descritas como atractivas, serian percibidas de manera más positiva (por ejemplo, más agradables, más dedicadas a los deportes y más heroicas), que las mujeres atletas que no fueron descritas de esta forma, ya que el hecho de ser atractivas "suaviza" la inconsistencia percibida del papel del género de una mujer atleta. Por otro lado, se esperaba que los hombres atletas, descritos como individuos atractivos, fueran percibidos de manera más negativa que los hombres a quienes no se describió de esa forma, ya que el esquema de género de los hombres atletas hace que las personas tengan la expectativa de que el tema central de un articulo de una revista sean sus habilidades atléticas y no su atractivo fisico.

Además, se predijo que los resultados estarían influidos por el género del participante. Puesto que las mujeres suelen ser más tolerantes hacia la inconsistencia de un esquema (por ejemplo, Fiala, Giuliano, Remlinger y Braithwaite, 1999; Greendorfer, 1993) y de las mujeres atletas, en general (Nixon, Maresca y Silverman, 1979), se predijeron interacciones de tres factores, como que los participantes masculinos percibirían la conducta tipica de género (es decir, artículos sobre mujeres atletas atractivas y hombres atletas

Él es un Laker 10

con habilidades atléticas) de forma positiva, mientras que las participantes mujeres tenderian a valorar el comportamiento atipico y no consistente con el papel genérico (es decir, artículos sobre mujeres atletas con habilidades atléticas u hombres atletas atractivos).

Método

Participantes

Los participantes fueron 92 estudiantes de licenciatura, predominantemente blancos (40 hombres, 52 mujeres), de una pequeña universidad en el sureste de Estados Unidos. Los participantes se reclutaron principalmente de cursos de introducción a la psicología y a la economía, y se les ofreció un crédito extra en sus cursos por completar el estudio. Otros participantes se obtuvieron de los equipos de hombres y mujeres de la División III de fútbol soccer en la misma universidad, a los cuales se les dieron pequeños premios como incentivos.

Diseño y materiales

Se utilizó un diseño entre sujetos 2 (género del atleta: mujer u hombre) x 2 (género del participante: mujer u hombre) x 2 (enfoque del articulo: atractivo fisico o habilidades atléticas), para explorar la forma en que se percibe a los atletas, en función de su género y del enfoque del articulo. Debido a que los artículos de periódicos y de revistas pocas veces *describen* únicamente los atributos físicos, se

La sección del método inicia inmediatamente después de la introducción (y no en una nueva página). La palabra "Método" se coloca en el centro, con letra

Los encabezados de las subsecciones, tales como "participantes", se escriben en cursivas, justificadas en el margen izquierdo y en un solo renglón.

Los autores determinaron que la subsección "Diseño y materiales" ayudaría a organizar el documento. No todos los trabajos deben incluir esta subsección, o podría haber una sección para el diseño y otra para los materiales. La meta principal es facilitar a lector la comprensión de la metodología.

incluyó en el articulo una fotografía del atleta hipotético. Para minimizar posibles confusiones, se realizó una prueba piloto para aparear los modelos masculino y femenino respecto a su atractivo físico y a su edad. De un grupo de 12 fotografías a color "de la cabeza", se seleccionó a un hombre (M atractivo = 6.47 en una escala de 7 puntos; *M* edad = 19.43) y a una mujer (M atractivo = 6.93; *M* edad = 21.20), debido a que recibieron puntuaciones similares en estos atributos. En el artículo que enfatizaba el atractivo físico del atleta, se le describía como "famoso(a), tanto por su cuerpo increíble, como por sus poderosos golpes", por ser una de las "cincuenta personas más atractivas en el mundo", según la revista *People Magazine*, y por haber firmado recientemente un contrato como modelo para realizar un calendario "húmedo y salvaje" para los trajes de baño *Speedo*, después de los Juegos Olímpicos. En contraste, en el articulo centrado en las habilidades atléticas, el atleta se describía como "famoso(a) por su increíble velocidad y poderosos golpes", por ser uno de los "cincuenta atletas más prometedores", según la revista *Sports Illustrated*, y por haber firmado recientemente un contrato como modelo para un calendario promocional de la marca *Speedo*.

Los símbolos estadísticos, como M *parala medida, se escriben en cursivas.*

Después de la lectura del articulo, se utilizó una hoja de respuestas diseñada para evaluar las reacciones de los participantes a los perfiles. En especifico, se pidió a los participantes que evaluaran, en escalas de 10 puntos que iban de 1

(nada) a 10 (muchísimo), diversas características del atleta, según percibían su feminidad (es decir, su ternura, alegría, afecto, simpatía, gentileza, compasión y calidez; alfa de Cronbach = 0.93), su respetabilidad (es decir, actitud de líder, respetabilidad, ambiciones, representatividad de Estados Unidos y buen modelo del rol de género; alfa de Cronbach = 0.90), habilidades atléticas (es decir, competitividad, talento, seriedad con el deporte y dedicación al deporte; alfa de Cronbach = 0.84) y su agresividad (es decir, dominancia y fuerza; alfa de Cronbach = 0.85). Por medio de la misma escala, los participantes también juzgaron otros aspectos del atleta, como qué tan agradable era, qué tan heroico, su similitud con el hombre o mujer "ideal", y su atractivo físico. Después se les dio la oportunidad de registrar respuestas abiertas sobre sus impresiones generales del atleta.

Además, puesto que los medios han argumentado que la cobertura consistente con el género es "lo que la gente quiere" (Kane, 1996), utilizamos una categoría secundaria de variables dependientes para evaluar el gusto del participante por el artículo hipotético. Por lo tanto, los participantes también calificaron (en la misma escala de 10 puntos) qué tanto les agradó el artículo (es decir, qué tanto halagaba el artículo al atleta, qué tan interesados estarían en leer más acerca del atleta y con cuánta precisión describía el artículo al atleta; alfa de Cronbach = 0.87). Después, los participantes dieron

Los extremos de las escalas (nada) o (muchísimo) se escriben en cursivas.

Este trabajo no utiliza encabezados de párrafo. Puede encontrar otros artículos en los que las subsecciones, como el procedimiento, se dividan, a su vez, por medio de encabezados de párrafo.

respuestas abiertas sobre su impresión general del artículo.

Se emplearon varios procedimientos para controlar las variables extrañas. Primero, puesto que lo apropiado del deporte para el género de un atleta es tan obvio, fue necesario elegir un deporte que no fuera especialmente percibido como masculino o femenino. Para determinar cuál deporte era considerado neutral, 15 participantes calificaron lo apropiado, de acuerdo a género, de 16 deportes, en una escala de 7 puntos que iba de 1 *(apropiado para hombres)* a 7 *(apropiado para mujeres).* En este estudio se eligió a la natación (M = 4.07), pues era un deporte percibido como neutral, en consistencia con investigaciones anteriores (Metheny, 1965). Después, para asegurar una representación balanceada del género en la muestra, se utilizó un diseño de bloques aleatorios, usados para asignar un número igual de participantes hombres y mujeres a cada una de las cuatro condiciones. Finalmente, para minimizar una mayor cantidad de varianza del error, la edad reportada del atleta (es decir, 21 años de edad) y el nombre (según la condición, *Nick Gleason o Nicole Gleason)* se mantuvieron constantes en todas las condiciones.

Procedimiento

Se abordó a participantes potenciales y se les dijo que el estudio era "una investigación sobre la percepción que tiene la gente de los atletas olímpicos". Una vez que estuvieron de acuerdo en

contestar el cuestionario, los participantes leyeron un relato hipotético de un atleta en un periódico (quien ostensiblemente habia competido en los Juegos Olímpicos de verano, de 1996), y después expresaron sus juicios en respuesta a la cobertura y al atleta involucrado. Todos los participantes recibieron perfiles idénticos, con excepción de los nombres de pila (es decir, el género) y el tipo de cobertura (es decir, enfocada en el atractivo o en las habilidades atléticas), que variaron según la condición experimental especifica.

Después de leer los cuatro perfiles (se incluyeron tres perfiles hipotéticos adicionales, como parte de otra investigación) y de completar las hojas de respuesta correspondientes, los participantes registraron sus respuestas a preguntas sobre datos demográficos (e.g., edad, género, nivel atlético y la cantidad de tiempo que dedican a seguir los deportes en los medios de comunicación), y otras medidas de personalidad, incluyendo el Inventario de Roles Sexuales de Bem (Bem, 1974) y la Escala de Igualdad del Rol Sexual (Beere, King, Beere y King, 1984).[1] Una vez finalizados los cuestionarios, se informó a los participantes que el articulo era hipotético, se les agradeció su participación y se fueron.

Resultados

Los datos se analizaron por medio de un análisis de varianza (ANOVA) entre sujetos 2

En general, evite el uso de las notas al pie de página. Si las utiliza, indíquelas por medio de un superíndice; el texto de la nota se escribe en una página separada al final del documento.

La sección de resultados se inicia inmediatamente después de la sección del método. La palabra "Resultados" se escribe con letra normal y se coloca en el centro.

Se indican las probabilidades exactas, aun cuando en ocasiones es apropiado el uso de "ns" para indicar que un resultado no es significativo.

Cuando se presenten los resultados de una prueba estadística, el nombre de la prueba se escribe en cursivas y va seguida por los grados de libertad en paréntesis. La p se refiere a la probabilidad de obtener el resultado, si la hipótesis nula es correcta.

(género del atleta: mujer u hombre) x 2 (género del participante: mujer u hombre) x 2 (enfoque del articulo: atractivo físico o habilidad atléticas).

Debido a que no se encontró un efecto del género del participante, los resultados reportados no incluyen esta variable. Se obtuvo una interacción entre el género del atleta x el enfoque del articulo, para las calificaciones de los participantes respecto al atractivo del atleta, F (1,83) = 5.56, p = 0.021. Como se esperaba, las mujeres atletas descritas en términos de su atractivo (M = 8.90) fueron consideradas más atractivas que aquellas descritas únicamente en términos de sus habilidades atléticas (M = 6.54; t(43) = 5.14, p = 0.0001); en contraste, no hubo diferencias en el atractivo percibido de los hombres atletas que fueron descritos como individuos atractivos (M = 8.18) o como individuos con habilidades atléticas (M = 7.46; t < 1, ns) , tal como se muestra en la figura 1. La magnitud del efecto r asociada con la comparación del atractivo percibido de las mujeres descritas como atractivas o con habilidades atléticas, fue de 0.617.

Cuando utilice figuras y tablas, debe mencionarlas en el texto.

Además, se obtuvieron varios efectos principales del enfoque del articulo. Los atletas cuya cobertura se enfocó en su atractivo, fueron considerados menos talentosos (M = 8.08), que los atletas que fueron descritos como poseedores de habilidades atléticas (M = 8.96), P(l,83) = 11.68, p = 0.001. Los atletas descritos como individuos atractivos también fueron considerados menos

agresivos (*M* = 6.22), que los atletas descritos con habilidades atléticas (*M* = 6.97), *F*(1,83) = 4.05, *p* = 0.047. También, los atletas representados como atractivos fueron considerados menos heroicos (M = 5.43), que los atletas descritos con habilidades atléticas (M = 6.52), F(1,83) = 5.74, p = 0.019. Finalmente, cuando el atractivo era el tema central del articulo, a las personas les gustó menos el articulo (M = 5.18), que cuando la cobertura se enfocó en las habilidades atléticas del atleta (M = 6.14), F(1,83) = 5.09, p = 0.027.

Discusión

La sección de discusión debe ir inmediatamente después de la sección de resultados. La palabra "discusión" se escribe en el centro, con letra regular.

Los resultados del presente estudio confirman que las percepciones de los atletas se ven influidas por el género del atleta y por el tipo de cobertura de los medios proporcionada en el articulo. De forma interesante, aun cuando se utilizó la misma fotografía en cada una de las condiciones, una mujer atleta, cuyo atractivo era el enfoque principal del articulo, era percibida con mayor atractivo fisico que una mujer atleta cuyos logros atléticos constituían el tema principal del articulo. Sin embargo, no se encontró el mismo patrón con los hombres atletas. Investigaciones previas han demostrado que las personas poseen esquemas más débiles respecto al atleta ideal que a la persona ideal, debido a que el público general tiene un menor número de experiencias (y, por lo tanto, menores asociaciones cognoscitivas) con la categoría muy

especifica de "un atleta ideal", en oposición a la categoría más amplia de "una persona ideal" (Martin y Martin, 1995). Se infiere que tal vez el esquema de una mujer atleta no es tan fuerte como el de un hombre atleta (debido a una menor "exposición") a través de su "aniquilación simbólica" por los medios; Gerbner, 1972) y, por lo tanto, la percepción que las personas tienen de las mujeres atletas es más maleable y abierta al cambio (especialmente en el atractivo). Por esto, este estudio implica que las personas son más capaces de confiar en información periférica (como el punto de vista proporcionada por el tipo de cobertura) para formar impresiones sobre una mujer atleta.

Sin embargo, sin importar el género del atleta, el enfoque en el atractivo, con la exclusión de la habilidad atlética, tuvo notables consecuencias en la forma en que fueron percibidos los atletas. De forma interesante, nuestros resultados indican que los hombres atletas también se ven afectados por coberturas que formen una imagen trivial, sin embargo, puesto que los hombres no suele ser descritos por los medios en términos de su atractivo físico (como las mujeres), esta cobertura prejuiciosa parece afectar de manera predominante a las mujeres atletas. Debido a este efecto negativo sobre las impresiones de mujeres y hombres atletas, los medios necesitan estar conscientes de: *a)* el daño que causa en la percepción de la gente el enfocarse en el atractivo de los atletas, *b)* el

hecho de que la gente podría preferir artículos que traten sobre las habilidades atléticas de un atleta, más que sobre su atractivo físico y c) la realidad de que ellos no son un simple reflejo de la opinión pública; sino que, de hecho, pueden moldearla de forma activa.

Es interesante el hecho de que el género de los participantes no haya sido un factor significativo en la calificación de los atletas o del articulo, hallazgo que se contrapone a algunos estudios previos (por ejemplo, Fisher, Genovese, Morris y Morris, 1977), pero que es consistente con otras investigaciones (por ejemplo, Michael, Gilroy y Sherman, 1984). Quizás esto refleje un cambio en las actitudes de los hombres hacia las mujeres atletas. A pesar de que las mujeres, por tradición, han aceptado más que los hombres el atletismo femenil (Nixon, Maresca y Silverman, 1979), con el éxito de las mujeres atletas en el ámbito profesional, olímpico, colegial y en la preparatoria, quizás los hombres ahora estén más interesados en los deportes femeninos y, por lo tanto, los aceptan más.

Un examen de las respuestas abiertas confirmó lo que los datos cuantitativos revelaron. Por ejemplo, una participante señaló inteligentemente el hecho de que la cobertura de la mujer atleta se centraba en su atractivo, "si yo fuera ella, me sentirla ofendida de que este articulo hable más de mi apariencia física que de mi talento, una actitud típica hacia las mujeres. Ellos no pueden resistirse a hablar de tu apariencia". Un hombre

participante hizo un señalamiento similar, "si esto fuera una edición de la revista *Cosmopolitan*, tal vez me habría gustado, pero no dice nada de la mujer como atleta". Incluso otro hombre participante dijo: "me pregunto cuáles son sus prioridades... ¿está ella utilizando "esto de nadar" para lograr una dulce carrera de modelo?". Las respuestas abiertas acerca de los hombres atletas descritos como atractivos, revelaron que también fueron percibidos de manera negativa. Una mujer participante escribió que este atleta era "un niño rico que es un buen nadador para que todos le digan lo maravilloso que es". Otra mujer participante señaló: "el articulo no menciona perfiles deportivos atléticos, sino únicamente la apariencia. Cuando se habla de un atleta de un deporte olímpico, esto es desconcertante".

Este estudio presentó varias limitaciones. Lo más notorio, el tipo de metodología utilizada (es decir, investigación de "efectos") no puede simular verdaderamente las consecuencias a largo plazo de los medios de comunicación, como una herramienta de socialización (Lewis, 1991). A pesar de que este tipo de investigación tiene ciertas limitaciones, otros investigadores han reconocido su potencial valor teórico y práctico. Por ejemplo, en su investigación sobre el *rostrismo*, Archer y sus colaboradores tomaron un hallazgo cualitativo (es decir, que las fotografías de mujeres en los medios generalmente incluyen el cuerpo completo, mientras que las fotografías de hombres suelen incluir únicamente

la cabeza) y determinaron de manera experimental, por medio de metodología cuantitativa, que este patrón da como resultado que los hombres parezcan ser más inteligentes y que las mujeres sean vistas como objetos (Archer, Iritani, Kimes y Barrios, 1983). De forma similar, la presente investigación ha mostrado las impresiones y efectos negativos que pueden surgir de los artículos que se enfocan exclusivamente en el atractivo fisico de los atletas. Es por medio de este proceso de triangulación y mediante el examen del mismo fenómeno desde distintos paradigmas y perspectivas disciplinarias, que empezaremos a ser capaces de comprender verdaderamente los patrones sociales y su efecto potencial sobre la sociedad.

Otra limitación potencial fue el hecho de que algunas de las condiciones en el estudio carecieron de validez externa (es decir, que los hombres atletas pocas veces son descritos en términos de su atractivo fisico). A pesar de que artículos como éstos pocas veces se encuentran en los medios impresos, fue importante incluirlos en el presente estudio, para poder determinar de forma empírica la manera en que las representaciones atípicas del género de hombres y mujeres atletas, afecta la percepción de la gente. Es decir, se tomó la decisión consciente de intercambiar validez interna por externa, en el presente caso. Finalmente, no se pidió a los participantes que reportaran su raza, por lo que este aspecto no fue tomado en cuenta en el análisis. Aun cuando la mayoría de los

participantes eran blancos, investigaciones anteriores indican que los participantes negros y blancos ponderan ciertas características de manera distinta cuando evalúan el atractivo de individuos blancos y negros (por ejemplo, Hebl y Heatherton, 1998) . Es probable que audiencias diferentes (en términos de la edad, el nivel socioeconómico, el nivel educativo, la orientación sexual y la orientación política) también presenten diferencias con respecto al gusto por los artículos que se enfocan en el atractivo físico y las habilidades atléticas de hombres y mujeres atletas, y sus posteriores impresiones de ellos.

Existen muchas oportunidades de investigaciones cuantitativas futuras en el área de las representaciones por género en los medios deportivos. Por ejemplo, las fotografías seleccionadas para este estudio incluían individuos blancos. Puesto que las personas tienen distintas expectativas y esquemas de las mujeres atletas negras, los resultados del presente estudio tal vez no puedan generalizarse a atletas de otras razas. Por ejemplo, por tradición es más aceptable que mujeres atletas de grupos minoritarios y de la clase trabajadora, participen en deportes inapropiados para su género (por ejemplo, básquetbol, fútbol soccer, hockey), que mujeres atletas blancas y de clase media, debido a que el primer grupo posee una percepción más dinámica de la feminidad (Cahn, 1994; Metheny, 1965). Por lo tanto, es necesario que se realicen más investigaciones para estudiar las

interacciones potenciales entre la raza del participante, el género del participante, la raza del atleta y el género del atleta (Gissendaner, 1994).

En un ámbito más amplio, deben realizarse más investigaciones experimentales cuantitativas para verificar de forma empírica lo que los estudios descriptivos cualitativos han estado reportando desde hace tiempo-que las mujeres atletas reciben una cobertura trivial por parte de los medios de comunicación. Por ejemplo, *las marcas de género* (es decir, la calificación de los equipos y las competencias atléticas femeniles a partir de las normas o estándares de las competencias masculinas) tienen gran prevalencia en las coberturas televisivas de mujeres atletas (Kane, 1996). Los escritores y comentaristas deportivos con frecuencia utilizan etiquetas genéricas para describir juegos en los que participan mujeres atletas (por ejemplo, "cuartos de final *de* mujeres"). Aunque las competencias atléticas de los hombres no se describen en términos de género (por ejemplo, "los cuartos de final" en lugar de "los cuartos de final de hombres"). Aun cuando los investigadores han especulado que este tipo de cobertura margina a las mujeres atletas, al hacerlas parecer "el otro" y no la norma, la investigación aún debe demostrar de forma empírica las consecuencias de las marcas de género. La exploración de éstos y otros tipos de representaciones del género estereotipadas, que afectan tanto a los hombres como las mujeres

atletas, es el siguiente paso importante en la literatura deportiva.

Mientras tanto, el presente estudio proporciona una perspectiva empírica para los florecientes campos psicológico y sociológico del estudio de los medios, los deportes y el género. En ninguna otra época de la historia las mujeres han tenido tanto ánimo personal (Weiss y Barber, 1995) o tantas oportunidades para participar en los deportes (Women's Sports Foundation, 1998) como ahora y, sin embargo, la cobertura de los deportes femeniles aún carece de la cobertura de los deportes masculinos, tanto en cantidad como en calidad. Los medios necesitan concientizarse de los efectos de sus coberturas, que producen imágenes triviales y marginales, y del hecho de que este tipo de cobertura tal vez no sea "lo que el público quiere", después de todo. Ojalá que, con un compromiso continuo y diligente por parte de los medios, los deportes serán considerados como una actividad aceptable y benéfica para las mujeres.

Las referencias se anotan en una nueva página. La palabra "referencias" se coloca en el centro de la parte superior de la página.

Cada referencia se inicia en un nuevo renglón y se maneja como un párrafo. El párrafo lleva un formato con "sangría colgante", en la que el primer renglón se coloca hacia el margen izquierdo y los renglones subsecuentes incluyen una sangría. Utilice la herramienta de la sangría colgante de su procesador de textos para hacer esto.

Los títulos de los libros y de las revistas se escriben en cursivas así como el número del volumen de las revistas.

El es un Laker 24

Referencias

Archer, D., Iritani, B., Kimes, D. D., y Barrios, M. (1983). Five studies of sex differences in facial prominence. *Journal of Personality and Social Psychology, 45,* 725-735.

Beere, C. A. King, D. W., Beere, D. B. y King, L. A. (1984). The Sex-Role Egalitarianism Scale: A measure of attitudes toward equality between the sexes. *Sex Roles, 10,* 563-576.

Bem, S. L. (1981). Gender schema theory: a cognitive account of sex typing. *Psychological Review, 88,* 354-364.

Bem, S. L. (1974). The measurement of Psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155-162.

Birrel, S. y Theberge, N. (1994). Ideological control on women in sport. En D. M. Costa y S. R. Guthrie (Eds.), *Women and sport: Interdisciplinary perpectives.* Champaign, IL: Human Kinetics.

Boutilier, M. A. y SanGiovanni, L. (1983). *The sporting woman.* Champaign, IL: Human Kinetics.

Burn, S. M., O'Neal, K. A. y Nederend, S. (1996). Childhood tomboyism and adult androgyny. *Sex Roles, 34,* 419-428.

Cahan, S. K. (1994). *Corning on strong: Gender and sexuality in twentieth-century women's sport.* New York: Free Press.

Coakley, J. J. y White, A. (1992). Making Decisions: Gender and sport participation

Estas referencias incluyen ejemplos de libros, libros editados, artículos de revistas científicas, artículos de la prensa popular y material de Internet. El Manual de Publicaciones de la APA *describe virtualmente todos los tipos de referencias que podría utilizar.*

among British adolescents. *Socíology of Sport Journal, 9,* 20-35.

Curtis, M. C. y Grant, C. H. B. (2001). *Gender equity in sports.* Recuperado el 15 de julio del 2001, de http://ww.bailiwick.lib. uiowa.edu/ge/

Daddario, G. (1998). *Women's sport and spectacle: gendered televisión coverage and the Olympic Games.* Westport, CT: Praeger.

Duncan, M. C. (1990). Sport photographs and sexual difference: Images of woman and men in the 1984 and 1988 Olympic Games. *Sociology of Sport Journal, 7,* 22-43.

Duncan, M. C. y Messner, M. A. (2000). *Gender stereotyping in televised sports: 1989, 1993 y 1999.* Los Angeles: Amateur Athletic Foundation of Los Angeles.

Eastman, S. T. y Billings, A. C. (1999). Gender parity in the Olympics: Hyping women athletes, favoring men athletes. *Journal of Sport and Social Issues, 23,* 140-170.

Fasting, K. (1999). Women and sport. *Olympic Review, 26,* 43-45.

Fiala, S. E., Giuliano, T. A., Remlinger, N. M. y Braithwaite, L. C. (1999). Lending a helping hand: The effects of gender stereo-types and gender on likelihood of helping. *Journal of Appliede Social Psychology, 29,* 2164-2176.

Fink, J. S. (1998). Female athletes and the media: Strides and stalemates. *Journal of Physical Education, Recreation & Dance, 69,* 37-45.

Fisher, A., Genovese, P., Morris, K. y Morris, H. (1977). Perceptions of females in sport. En D. Landers y R. Christina (Eds.), *Psychology of motor behavior and sport* (pp. 96-118). Champaign, IL: Human Kinetics.

Fiske, S. T. y Taylor, S. E. (1991). *Social cognitlon* (2ª ed.). New York: McGraw-Hill.

Gerbner, G. (1972). Violence in televisión drama: Trends and symbolic functions. En G. A. Comstock y I. A: Rubinstein (Eds.), *Media content and control: Televisión and social behavior, Vol. 1,* (pp. 28-187). Washington, DC: U.S. Government Printing Office.

Gissendaner, C. H. (1994). African-american women and competitive sport, 1920-1960. En S. Birrel y C. L. Colé (Eds.), *Women, sport, and culture* (pp. 81-92). Champaign, IL: Human Kinetics.

Greendorfer, S. L. (1993). Gender role stereotypes and early childhood socialization. *Psychology of Women Quarterly, 18,* 85-104.

Hebl, M. R. y Heatherton, T. F. (1998). The stigma of obesity in women: The difference is black and white. *Personality and Social Psychology Bulletin, 24* (4), 417-426.

Higgs, C. T. y Weiller, K. H. (1994). Gender bias and the 1992 Summer Olympic Games: An analysis of televisión coverage. *Journal of Sport and Social Issues, 3,* 234-246.

Hilliard, D. C. (1984). Media images of male and female professional athletes: An interpretative analysis of magazine articles. *Sociology of Sport Journal, 1,* 251-262.

Kane, M. J. (1996). Media coverage of the post Title IX female athlete. *Duke Journal of Gender Law & Policy, 3,* 95-127.

Kane, M. J. y Parks, J. B. (1992). The social construction of gender difference and hierarchy in sport journalism-Few new twists on very oíd themes. *Women in Sport and Physical Activity Journal, 1,* 49-8 3.

Knight, J. L., Giuliano, T. A. y Sanchez-Ross, M. G. (2001). Famous or infamous? The influence of celebrity status and race on perceptions of responsibility for rape. *Basic and Applied Social Psychology, 23,* 783-190.

Lewis, J. (1991). *The ideological octopus*. New York: Routledge.

Lumpkin, A. y Williams, L. D. (1991). An analysis of *Sports Illustrated* feature articles, 1954-1987. *Sociology of Sport Journal, 8,* 1-15.

Martin, B. A. y Martin, J. H. (1995). Compared perceived sex role orientation of the ideal male and female athlete to the ideal male and

female person. *Journal of Sport Behavior, 18,* 286-301.

Martin, E. (1987). *The woman in the body.* Boston: Beacon Press.

Messner, M. (1988). Sports and male domination: The female athlete as contested ideological terrain. *Sociology of Sport Journal,* 5, 19 7-211.

Metheny, E. (1965). Symbolic forms of movement: The feminine image in sports. En E. Metheny (Ed.), *Connotations of movement in sport and dance* (pp. 43-56). Dubuque, IA:Wm. C. Brown.

Michael, M. E., Gilroy, F. D. y Sherman, M. F. (1984). Athletic similarity and attitudes toward woman as factors in the perceived physical attractiveness and liking of a female varsity athlete. *Perceptual and Motor Skills, 59,* 511-518.

Nixon, H. L., Maresca, P. J. y Silverman, M .A. (1979). Sex differences in college students'acceptance of female into sport. *Adolescence, 14,* 755-764.

Rosenthal, R. y Jacobson, L. F. (1968). *Pygmalion in the classroom: Teacher expectations and pupils' intellectual development.* New York: Holt, Rinehart & Winston.

Sabo, D. y Jansen, S. C. (1992). Images of men in sport media: The social reproduction of gen-der order. *Masculinity and the media.* Newbury Park, CA: Sage.

Sagas, M. , Cunningham, G. B., Wigley, B. J. y Ashley, F. B. (2000). Internet coverage of university softball and baseball websites: The inquiry continúes. *Sociology and Sport Journal,* 17(2), 198-212.

Salwen, M. B. y Wood, N. (1994). Depictions of female athletes on *Sports Illustrated* covers, 1957-89. *Journal of Sport Behavior, 17,* 98-107.

Tuggle, C. A. y Owen, A. (1999). A descriptive analysis of NBC's coverage of the Centennial Olympics. *Journal of Sport and Social Issues, 23* (2), 171-182.

Weiss, M. R. y Barber, H. (1995). Socialization influences of college female athletes: A tale of two decades. *Sex Roles, 33,* 129-140.

The Women's Sport Foundation Report: Sport and Teen Pregnancy. (Mayo, 1998). East Meadow, NY: The Women's Sport Foundation.

Woodcock, A. T. (1995). Media coverage of boys' and girls' high school ice hockey in Minnesota. *Melpomene Journal, 14,* 27-29.

Wulf, S. (August 4, 1997). The NBA's sister act:Pro hoops gets a new season and a new gender with the successful start-up of the WNBA. *Time,* 41-43.

Él es un Laker 30

Nota del autor

Jennifer L. Knight y Traci A. Giuliano, Departamento de Psicología, Southwestern University.

Jennifer L. Knight se encuentra ahora en el Departamento de Psicología en Rice University.

Agradecemos a Dan Hilliard y Jacqueline Muir-Broaddus por su ayuda en la fase del diseño del proyecto, y a Marie Helweg-Larsen por sus útiles comentarios a un borrador anterior de este manuscrito. También queremos dar las gracias a tres revisores anónimos por sus útiles comentarios y sugerencias. Una versión de este articulo se presentó en el 12° Congreso anual de la American Psychological Society, en Miami, en el 2000.

La correspondencia referente a este articulo debe enviarse a Jennifer Knight, Departamento de Psicología, MS-25, Rice University, P. 0. Box 1892, Houston, TX 77251-1892. E-mail: jknight@rice.edu o giuliant@southwestern.edu

La nota del autor se escribe en una página separada.

El primer párrafo proporciona la afiliación institucional del autor o autores. El segundo párrafo incluye cualquier cambio en la afiliación de los autores. El párrafo final proporciona información para establecer contacto. La nota del autor debe limitarse a cuatro párrafos.

Algunas revistas científicas pueden requerir que la nota del autor se incluya en la página del título.

Si se utilizan notas al pie de página, se escriben en una página separada.

Él es un Laker 31

Nota al pie de página

[1] Estas variables demográficas y de personalidad (por ejemplo, nivel atlético, edad, BSRI [Bem, 1974] y SRES [Beere *et al.*, 1984]) se incluyeron para explorar los potenciales efectos moderadores que podrian tener sobre las interacciones y los efectos principales. Sin embargo, análisis preliminares no revelaron ningún patrón significativo de los resultados, y por ello no se comenta más de estas variables.

Él es un Laker 32

Texto de la Figura

Figura 1. Atractivo percibido del atleta en función de su género y el enfoque del articulo (atractivo o habilidad atlética).

Escriba el pie o texto de la figura en una nueva página. Todos los textos de figuras se anotarán en esta página.

El número de la figura se escribe en cursivas. Sólo la primera palabra del texto de la figura se escribe con mayúscula inicial.

Es probable que en los reportes de estudiantes no requieran una página con el texto de las figuras, sino que éste puede incluirse en la página que contiene las figuras (como sucede en un artículo publicado).

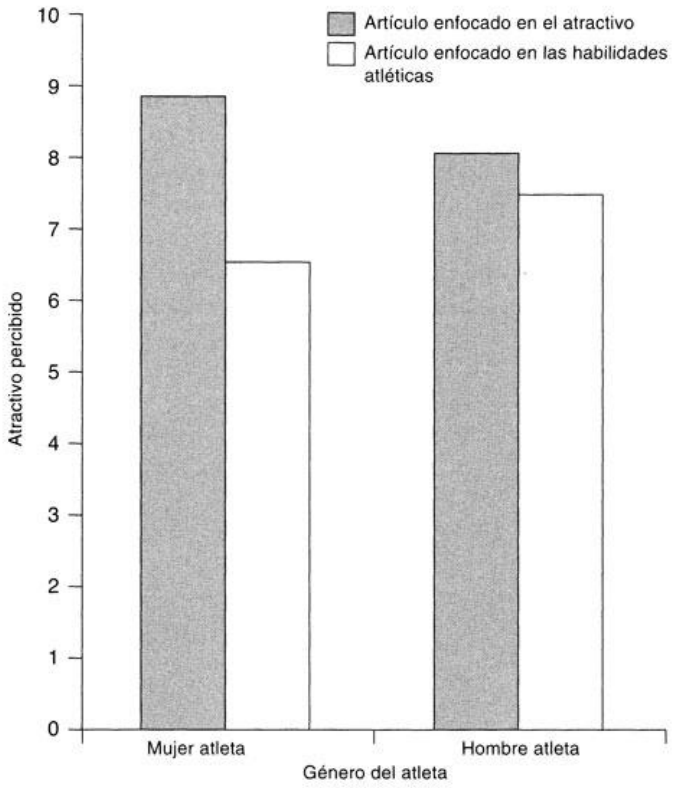
Cada figura debe ir en una página se

La página con la figura no incluye encabezado o número de página. Para identificar la figura, escriba el número a lápiz en el reverso de la página.

Incluya todas las etiquetas necesarias para interpretar la figura.

Utilice una fuente Sans serif como Arial; el tamaño máximo de la fuente es de 14 puntos.

Observe que, en una gráfica lineal o de barras, la variable independiente se coloca en el eje horizontal y la variable dependiente en el eje vertical.



Apéndice B
Pruebas estadísticas

El propósito de este apéndice es el de proporcionar las fórmulas y los procedimientos de cálculo para el análisis de datos. No se incluyen todas las pruebas estadísticas, aunque sí una variedad que puede ser apropiada para muchos de los diseños de investigación que usted podría utilizar.

Examinaremos algunos estadísticos descriptivos e inferenciales. Sin embargo, antes de estudiar la estadística, debe repasar las propiedades de las escalas de medición descritas en el capítulo 5. Recuerde que existen cuatro tipos de escalas de medición: nominal, ordinal, de intervalo y de razón. Las escalas nominales no poseen propiedades numéricas, las escalas ordinales únicamente proporcionan información de orden, y las de intervalo y de razón tiene intervalos iguales entre los puntos de la escala. Además, las escalas de razón incluyen un cero absoluto. En el capítulo 13 aprendimos que el análisis estadístico apropiado se determina por el tipo de diseño y de escala de medición, empleados en el estudio. Conforme avancemos, la discusión de las diversas pruebas estadísticas le indicarán las restricciones pertinentes en las escalas de medición.

Los ejemplos que presentamos incluyen sencillos conjuntos de datos, de modo que los cálculos puedan realizarse con facilidad a mano, con el uso de una calculadora. Sin embargo, se dará cuenta de que los cálculos son tediosos y de las posibilidades de cometer errores cuando se trabaja con grandes conjuntos de datos y cuando se necesita realizar muchos análisis estadísticos en un estudio. Existen programas de cómputo para hacer análisis estadísticos, que facilitan el proceso y reducen los errores de cálculo.

ESTADÍSTICA DESCRIPTIVA

Con el conocimiento de los tipos de escalas de medición, podemos iniciar el estudio de las técnicas estadísticas. Podemos empezar con dos formas de describir un conjunto de puntajes: la tendencia central y la variabilidad.

Medidas de tendencia central

Una medida de tendencia central proporciona un número que describe los puntajes de un grupo entero, o en promedio. Disponemos de 3 distintas medidas de tendencia central: la moda, la mediana y la media.

La moda La moda es el puntaje que ocurre con mayor frecuencia. La tabla B.l muestra un conjunto de puntajes y los estadísticos descriptivos que estudiamos en esta sección. El puntaje más frecuente en estos datos es 5; no es necesario hacer cálculos para encontrar la moda. La moda puede emplearse con cualquiera de los cuatro tipos de escalas de medición. Sin embargo, es la única medida de tendencia central que puede emplearse con datos de escala nominal. Si usted está midiendo el género y encuentra que hay 100 mujeres y 50 hombres, la moda es "mujer", puesto que es la categoría que ocurre con mayor frecuencia en la escala nominal.

La mediana La mediana es el puntaje que divide al grupo en dos mitades: el 50 por ciento de los puntajes están por debajo de la mediana y el 50 por ciento por arriba de ella. Cuando los puntajes están ordenados del más bajo al más alto (como en la tabla B.l), es fácil calcular la mediana. Si hay un número impar de puntajes, simplemente localice el puntaje de en medio (por ejemplo, si hay 11 puntajes, el sexto es la mediana, porque hay 5 por arriba y 5 por debajo). Si tenemos un número par de puntajes, la mediana es el punto medio entre los dos puntajes de en medio. En los datos de la tabla B.l hay 10 puntajes, por lo que el quinto y sexto son los que se encuentran a la mitad. Para calcular la mediana, sumamos los dos puntajes de la mitad y los dividimos entre 2. Así, en la tabla B.l, la mediana es

$$\frac{5 + 5}{2} = 5$$

La mediana puede utilizarse con datos ordinales, de intervalo y de razón. Sin embargo, lo más común es que se utilice con datos ordinales, debido a que el cálculo de la mediana toma en cuenta únicamente el orden de los puntajes y no su magnitud real.

La media La media sí toma en cuenta la magnitud real de los puntajes. Por lo tanto, se basa en más información acerca de los puntajes que la moda y la mediana. Sin embargo, únicamente es apropiada para datos de intervalo o de razón. La media es la suma de los puntajes de un grupo, dividida entre el número de puntajes. La fórmula para calcular la media se puede expresar como

$$\overline{X} = \frac{\Sigma X}{N}$$

donde X es el símbolo de la media. En esta fórmula, X representa el puntaje obtenido por un individuo, y el símbolo X indica que los puntajes deben sumarse. El símbolo IX puede leerse como "sumatoria de las X^y y sólo es una indicación de que los puntajes deben sumarse. Por consiguiente, XX de los datos de la tabla B.l es

$$1 + 2 + 4 + 4 + 5 + 5 + 5 + 6 + 6 + 7 = 45$$

TABLA B.1 Estadísticos descriptivos de un conjunto de puntajes

Puntajes	Estadísticos descriptivos
1	Moda = 5
2	
4	Mediana = 5
4	
5	$\bar{X} = \frac{\Sigma X}{N} = 4.5$
5	
5	
6	Rango = 6
6	
7	$s^2 = \frac{\Sigma(X - \bar{X})^2}{N-1} = \frac{\Sigma X^2 - N\bar{X}^2}{N-1} = \frac{233 - 202.5}{9} = 3.388$
$\Sigma X = 45$	
$\Sigma X^2 = 233$	$s = \sqrt{s^2} = 1.84$
$N = 10$	

La *N* en la fórmula simboliza el número de puntajes en el grupo. En nuestro ejemplo, $N = 10$, y ahora podemos calcular la media:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{45}{10} = 4.5$$

Medidas de variabilidad

Además de describir la tendencia central del conjunto de puntajes, debemos describir qué tanto varían entre ellos. ¿Cuánta dispersión hay en el conjunto de puntajes?

El rango El rango es el puntaje más alto, menos el puntaje más bajo. En nuestro ejemplo, el rango es 6. Sin embargo, el rango no es un estadístico muy útil debido a que se basa únicamente en dos puntajes de la distribución. No toma en cuenta toda la información disponible en el conjunto completo de puntajes.

La varianza y la desviación estándar La varianza, y un estadístico relacionado llamado desviación estándar, utiliza todos los puntajes para producir una medida de variabilidad. La varianza indica el grado en que los puntajes varían respecto a la media grupal. La fórmula para la varianza (que se simboliza s^2) es

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{N-1}$$

donde $(X- X)^2$ es un puntaje individual, X, menos la media, X, elevado después al cuadrado. Así, $(X-X)^2$ es la desviación cuadrada de cada puntaje, respecto a la media. El símbolo Z indica que estos puntajes de desviación cuadrada se suman. Finalmente, la división entre N- 1 produce la media de las desviaciones cuadradas. La varianza, entonces, es la media de las desviaciones cuadradas, a partir de la media grupal. Se utilizan las desviaciones cuadradas porque las desviaciones solas sumarían cero. En la mayoría de los casos, se utiliza TV- 1, con propósitos estadísticos, debido a que los puntajes representan una muestra y no a la población completa. Conforme el tamaño de la muestra aumenta, no importa si se emplea N o N - 1.

Podemos utilizar los datos de la tabla B.l para ilustrar el cálculo de la varianza. $H(X - X)^2$ es igual a

$$(1 - 4.5)^2 + (2 - 4.5)^2 + (4 - 4.5)^2 + (4 - 4.5)^2 + (5 - 4.5)^2 + (5 - 4.5)^2 + (5 - 4.5)^2 + (6 - 4.5)^2 + (6 - 4.5)^2 + (7 - 4.5)^2 - 30.50$$

El siguiente paso es dividir $S(X - X)^2$ entre N - 1. Entonces, el cálculo de la varianza es

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{N - 1} = \frac{30.50}{9} = 3.388$$

Una fórmula más sencilla para el cálculo de la varianza es

$$s^2 = \frac{\Sigma X^2 - N\bar{X}^2}{N - 1}$$

donde IX^2 es la sumatoria del cuadrado de los puntajes individuales, y X^2 es la media cuadrada. Usted puede confirmar que las dos fórmulas son idénticas si calcula la varianza por medio de la fórmula más sencilla (recuerde que IX^2 le indica que eleve al cuadrado cada puntaje y que después sume los puntajes cuadrados). Es más fácil trabajar con la fórmula sencilla cuando se tienen demasiados datos, ya que no es necesario calcular cada desviación.

La desviación estándar es la raíz cuadrada de la varianza. Dado que la varianza utiliza puntajes al cuadrado, no describe la cantidad de variabilidad en las mismas unidades de medición que la escala original. La desviación estándar (s) corrige este problema. Así, la desviación estándar es el promedio de la desviación de los puntajes respecto a la media.

PRUEBAS DE SIGNIFICACIÓN ESTADÍSTICA

Esta sección describe varias pruebas de significación estadística. Estas pruebas se utilizan para determinar la probabilidad de que los resultados de la inves-

tigación se deban al error aleatorio. Todas emplean la lógica de la hipótesis nula, que estudiamos en el capítulo 13. En esta sección abordaremos dos pruebas de significación: la chi cuadrada y el análisis de varianza o prueba *F*.

Chi cuadrada (X^2)

La prueba chi cuadrada (la letra griega chi, al cuadrado) se aplica a datos de escala nominal. Se utiliza cuando los datos son frecuencias, es decir, el número de sujetos que caen en cada una de diferentes categorías.

La chi cuadrada se puede emplear con el método experimental o con el correlacional. Se usa junto con el método experimental cuando la variable dependiente se mide en una escala nominal, y con el método correlacional cuando ambas variables se miden en escala nominal.

Ejemplo Suponga que desea saber si existe una relación entre el género y la mano dominante. Para esto, obtiene una muestra de 50 hombres y de 50 mujeres y les pregunta si son diestros, zurdos o ambidiestros (que emplean ambas manos con la misma habilidad). Su grupo de datos implica la clasificación de cada persona como hombre o mujer y como diestro, zurdo o ambidiestro.

En la tabla B.2 se presentan datos ficticios de un estudio de este tipo. Las frecuencias con la etiqueta "O", en cada una de las seis celdas de la tabla, se refiere al número de hombres y mujeres que caen en cada una de las tres categorías de la mano dominante. Las frecuencias con la etiqueta "E" se refieren a las frecuencias esperadas, si la hipótesis nula es correcta. Es importante que cada sujeto esté en una sola de las celdas cuando se emplea la chi cuadrada (es decir, que ningún sujeto puede considerarse hombre y mujer o zurdo y diestro).

La prueba de chi cuadrada examina la magnitud en que las frecuencias observadas en realidad en el estudio, difieren de las frecuencias esperadas, si la hipótesis nula es correcta. La hipótesis nula establece que no existe relación alguna entre el género y la mano dominante; que los hombres y las mujeres no difieren en esta característica.

La fórmula para el cálculo de la chi cuadrada es

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

donde O es la frecuencia *observada* en cada celda, E es la frecuencia *esperada* en cada celda y el símbolo Z se refiere a la sumatoria de todas las celdas. Los pasos para calcular el valor de $\%^2$ son:

Paso 1: ordene las frecuencias observadas, como en la tabla B.2. Observe que además de las frecuencias observadas en cada celda, la tabla incluye totales por renglón, totales por columna y el número total de observaciones *(N).*

Paso 2: calcule las frecuencias esperadas para cada celda en la tabla. La fórmula para determinar la frecuencia esperada es

$$E = \frac{\text{total de renglón} \times \text{total de columna}}{N}$$

donde el total de renglón se refiere al total de renglón para la celda y el total de columna se refiere al total de columna para la celda. Por lo tanto, la frecuencia esperada de la celda 1 (hombre diestro) es

$$E_1 = \frac{50 \times 50}{100} = 25$$

Las frecuencias esperadas de cada celda se muestran en la tabla B.2, debajo de las frecuencias observadas.

TABLA B.2 Datos de un estudio hipotético sobre la mano dominante: prueba chi cuadrada

Género del sujeto	Mano dominante			Totales de renglón
	Diestro	Zurdo	Ambidiestro	
Hombre	$O_1 = 15$ $E_1 = 25$	$O_2 = 30$ E_2 $E_2 = 20$	$0_3 = 5$ $E_3 = 5$	50
Mujer	$0_4 = 35$ $E_4 = 25$	$O_5 = 10$ $E_5 = 20$	$0_6 = 5$ E_6-5	50
Totales de columna	50	40	10	$N = 100$

Cálculos:

Número de celda	$\frac{(O-E)^2}{E}$
1	4.00
2	5.00
3	0.00
4	4.00
5	5.00
6	0.00
	$\Sigma = 18.00$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 18.00$$

Paso 3: calcule $(O - E)^2/E$ para cada celda. Para la celda 1, el cálculo es

$$\frac{(15-25)^2}{25} = \frac{100}{25} = 4.00$$

Paso 4: obtenga el valor de X^2 al sumar los valores de $(O - E)^2/E$, calculados en el paso 3. Los cálculos para obtener X^2 en el ejemplo, se muestran en la tabla R.2.

Significación de chi cuadrada La significación del valor obtenido de X^2 puede evaluarse consultando la tabla de valores críticos de X^2. En la tabla C.2, del apéndice C, se incluye una tabla de valores críticos de X^2. Los valores críticos indican el valor al que la X^2 obtenida debe ser igual o mayor para ser significativa al nivel 0.10, al nivel 0.05 y al nivel 0.01.

Para poder emplear la tabla de valores críticos de X^2, así como otras tablas estadísticas, debe entender el concepto *de grados de libertad (gl).* El valor crítico de X^2 en cualquier estudio, depende de los grados de libertad. Los grados de libertad se refieren al número de puntajes que pueden variar libremente. En la tabla de categorías de una prueba chi cuadrada, el número de grados de libertad es el número de celdas en las que las frecuencias pueden variar libremente, una vez que conocemos los totales de los renglones y los totales de las columnas. Los grados de libertad de la chi cuadrada se calculan fácilmente:

$$g\,l = (R - l)(C - l)$$

donde R es el número de renglones en la tabla y C es el número de columnas. En nuestro ejemplo de la tabla B.2, hay dos renglones y tres columnas, por lo que existen 2 grados de libertad. En un estudio que incluya tres renglones y tres columnas habrían 4 grados de libertad, y así sucesivamente.

Para utilizar la tabla C.2, obtenga los grados de libertad correctos y luego determine el valor crítico de X^2, necesario para rechazar la hipótesis nula al nivel de significación elegido. Con 2 grados de libertad, el valor obtenido de X^2 debe ser *igual o mayor que* el valor crítico de 5.991, para ser significativo al nivel 0.05. Existe solamente un 0.05 de probabilidad de obtener una X^2 de 5.991, si ocurre error aleatorio. Puesto que la X^2 obtenida en nuestro ejemplo es 18.00, podemos rechazar la hipótesis nula que establece que no existe una relación entre el género y la mano dominante. (La chi cuadrada está basada en datos ficticios, pero sería relativamente fácil determinar si en realidad existe dicha relación).

Comentarios finales La prueba chi cuadrada es sumamente útil y se emplea con frecuencia en todas las ciencias del comportamiento. La fórmula descrita para calcularla se aplica a estudios en los que hay más categorías en cualquiera de las variables. Sin embargo, una nota precautoria: cuando ambas variables tienen únicamente dos categorías, de modo que sólo hay dos renglones y dos columnas, la fórmula cambia ligeramente. En tal caso, la fórmula es

$$\chi^2 = \Sigma \frac{(\{|O - E|\} - 0.5)^2}{E}$$

donde | O - E | es el valor absoluto de O - E y 0.5 es una constante que se resta de cada celda.

Análisis de varianza (Prueba F)

El análisis de varianza, o prueba *F,* se utiliza para determinar si existe una diferencia significativa entre grupos medidos en una escala de intervalo o de razón. Los grupos pueden haberse formado por medio del método experimental o el correlaciona!; lo importante es que al menos se haya utilizado una medida de intervalo. El análisis de varianza puede usarse con diseños de grupos independientes o diseños de medidas repetidas. A continuación presentamos los cálculos para obtener la *F* en ambos tipos de diseños.

Análisis de varianza: una variable independiente

Para ilustrar las aplicaciones del análisis de varianza, considere un experimento hipotético sobre la distancia física y la revelación de información. Usted cree que las personas revelarán más información de sí mismos a un entrevistador, cuando estén sentadas más cerca de él que cuando estén más lejos. Para probar esta idea, usted lleva a cabo un experimento con entrevistas. A los participantes se les dice que se están estudiando técnicas de entrevista. Cada uno de los participantes se sienta en una habitación; el entrevistador entra a la habitación y se sienta a una de tres distancias del participante: cerca (0.60 metros), intermedia (1.20 metros) o lejos (1.80 metros). La distancia elegida por el entrevistador es la manipulación de la variable independiente. Se asigna aleatoriamente a los participantes a las tres condiciones de distancia y el comportamiento del entrevistador se mantiene constante en todas las condiciones. La entrevista consiste de unas preguntas, y la variable dependiente es el número de aseveraciones con contenido personal de los participantes, durante la entrevista.

En la tabla B.3 se muestran datos ficticios para este experimento. Observe que se trata de un diseño de grupos independientes, con cinco sujetos en cada grupo. Los cálculos de la varianza sistemática y de la varianza de error implican la obtención de la *suma de cuadrados* de los distintos tipos de varianza.

Suma de cuadrados La suma de cuadrados se refiere a la *suma de las desviaciones cuadradas a partir de la media.* El cálculo de un análisis de varianza para los datos de la tabla B.3 incluye tres sumas de cuadrados: 1) SC_{TOTAL} la suma de las desviaciones cuadradas del puntaje de cada individuo, a partir de la gran media; 2) SC_A la suma de las desviaciones cuadradas de cada una de las medias grupales, a partir de la gran media y 3) SC_{ERROR}, la suma de las desviaciones cuadradas de los puntajes individuales, a partir de su respectiva media grupal. La "A" en SC_A se utiliza para indicar que nos referimos a la varianza sistemática asociada con la variable independiente *A.*

Las tres sumas de cuadrados son desviaciones respecto a una media (recuerde que calculamos estas desviaciones anteriormente, cuando hablamos de la varianza de un conjunto de puntajes). Podríamos calcular las desviaciones de manera directa con los datos de la tabla B.3, pero estos cálculos son difíciles, por lo que utilizaremos fórmulas abreviadas para los cálculos. Las fórmulas son

$$E = \frac{\text{total de renglón} \times \text{total de columna}}{N}$$

Podrá notar aquí que $SC_{TOTAL} = SC_A + SC_{ERROR}$. Los cálculos reales se presentan en la tabla B.3.

$$E_1 = \frac{50 \times 50}{100} = 25$$

TABLA B.3 Datos de un experimento hipotético sobre la distancia y la revelación de información personal: análisis de varianza

Género del sujeto	Mano dominante: Diestro	Zurdo	Ambidiestro	Totales de renglón
Hombre	$O_1 = 15$ $E_1 = 25$	$O_2 = 30$ $E_2 = 20$	$O_3 = 5$ $E_3 = 5$	50
Mujer	$O_4 = 35$ $E_4 = 25$	$O_5 = 10$ $E_5 = 20$	$O_6 = 5$ $E_6 = 5$	50
Totales de columna	50	40	10	$N = 100$

Cálculos:

Número de celda	$\frac{(O-E)^2}{E}$
1	4.00
2	5.00
3	0.00
4	4.00
5	5.00
6	0.00
	$\Sigma = 18.00$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 18.00$$

Paso 3: calcule $(O - E)^2/E$ para cada celda. Para la celda 1, el cálculo es

$$\frac{(15-25)^2}{25} = \frac{100}{25} = 4.00$$

La $\sum X^2$ es la suma de los puntajes cuadrados de todos los sujetos del experimento. Primero se eleva al cuadrado cada puntaje, y después se suman los resultados. Así, para los datos de la tabla B.3, $\sum X^2$ es $33^2 + 24^2 + 31^2$, y así sucesivamente, hasta haber elevado al cuadrado todos los puntajes y luego sumar los resultados. Si usted realiza los cálculos a mano o con una calculadora de bolsillo, lo más conveniente sería obtener $\sum X^2$ de los puntajes de cada grupo *y* después sumarlos para el cálculo final. Esto se hizo con los datos de la tabla. La G se refiere al gran total de todos los puntajes. Esto implica sumar los puntajes de todos los sujetos; después se eleva el gran total al cuadrado y el resultado se divide entre N, el número total de sujetos en el experimento. Al calcular la suma de cuadrados, manténgalos siempre bien etiquetados, ya que puede simplificar cálculos posteriores al recurrir a éstos. Una vez que ha obtenido SC_{TOTAL} puede calcular SC_A.

SC_A La fórmula de la SC_A es

$$\Sigma \frac{T_a^2}{n_a} - \frac{G^2}{N}$$

La T_a en esta fórmula se refiere al total de los puntajes en el Grupo *a* de la variable independiente *A* (T_a es una abreviatura de $\sum X$ de cada grupo. [Recuerde el cálculo de $\sum X$ cuando estudiamos la media.] El símbolo T_a se utiliza para evitar el uso de demasiados signos de $\sum$ en los procedimientos de cálculo). La *a* simboliza el número específico del grupo; entonces T_a es un símbolo general de T_1 T_2 y T_3. Si observamos los datos de la tabla B.3, T_1 = 151, T_2= 118 y T_3 = 72. Estas son las sumas de los puntajes de cada grupo. Después de obtener T_a, calculamos elevando T_a al cuadrado. Ahora dividimos T^2 entre n_a, el número de sujetos en el grupo *a*. Una vez que hemos calculado T^2_a/n_a para cada grupo, se suman las cantidades, como lo indica el símbolo E.

Note que la segunda parte de la fórmula, G^2/N, se obtuvo con el cálculo de SC_{TOTAL}. Puesto que ya tenemos esta cantidad, no es necesario volver a obtenerla para el cálculo de SC_A Después de obtener SC_A' podemos calcular SC_{ERROR}.

SC_{ERROR} La fórmula para la SC_{ERROR} es

$$\Sigma X^2 - \Sigma \frac{T_a^2}{n_a}$$

Ambos valores se calcularon antes, al obtener SC_{TOTAl} y SC_A. Para obtener $SC_{ERROR)}$ únicamente encuentre estas cifras y realice la resta respectiva.

Para verificar los cálculos, nos aseguramos de que $SC_{TOTAL} = SC_A + SC_{ERROR}$.

El siguiente paso en el cálculo del análisis de varianza es la obtención de los *cuadrados medios* de cada una de las sumas de cuadrados. Después podremos calcular el valor de *F*. Los cálculos necesarios se muestran en la tabla B.4, que incluye un resumen del análisis de varianza. La creación de una tabla de resumen es la forma más sencilla de completar los cálculos.

Cuadrados medios Después de obtener las sumas de cuadrados, es necesario calcular los cuadrados medios. Cuadrado medio se refiere a la *media de la suma de las desviaciones cuadradas respecto a la media,* o en términos más sencillos, la media de la suma de los cuadrados. El cuadrado medio (CAÍ) es la suma de los cuadrados, dividida entre los grados de libertad. Los grados de libertad se determinan por el número de puntajes en la suma de cuadrados que pueden variar libremente. Los cuadrados medios son las varianzas que se utilizan para calcular el valor de *F*.

En la tabla B.4 puede ver que los cuadrados medios que requerimos son el cuadrado medio de *A* (varianza sistemática) y el cuadrado medio del error (varianza de error). Las fórmulas son

$$CM_A = SC_A / gl_A$$
$$CM_{ERROR} = SC_{ERROR} / gl_{ERROR}$$

donde $gl_A = a - 1$ (el número de grupos menos uno) *y* $gl_{ERR0R} = N - a$ (el número total de sujetos menos el número de grupos).

Obtención del valor *F* El valor de *F* se obtiene al dividir CM_A entre CM_{ERR0R}. Si sólo opera error aleatorio, el valor esperado de *F* es 1.0. A mayor magnitud del valor $F_{}$ menor es la probabilidad de que los resultados del experimento se deban al error aleatorio.

Significación de *F* Para determinar la significación del valor *F* obtenido, es necesario compararlo con un valor crítico de *F*. La tabla C.4, en el apéndice C, incluye los valores críticos de *F* para los niveles 0.10, 0.05 y 0.01 de significación. Para encontrar el valor crítico de i[7], localice en la tabla los grados de libertad del numerador del cociente (la varianza sistemática) y los grados de libertad del denominador del cociente *F* (la varianza del error). La intersección de ambos grados de libertad en la tabla corresponde al valor crítico de *F*.

Para esta muestra de datos, los grados de libertad correspondientes son 2 y 12 (véase la tabla B.4). El valor crítico de *F* de la tabla C.4 es 3.89, a un nivel de significación de 0.05. Para que los resultados obtenidos sean significativos, el valor *F* calculado debe ser igual o mayor que el valor crítico. Dado que el valor *F* obtenido en la tabla B.4 (23.15) es mayor que el valor crítico, concluimos que los resultados son significativos y rechazamos la hipótesis nula que afirma que las medias de los grupos son iguales en la población.

Comentarios finales El análisis de varianza para una variable independiente, con un diseño de grupos independientes, puede emplearse cuando hay dos o más grupos en el experimento. Las fórmulas generales proporcionadas son apropiadas para todo este tipo de diseños. Además, sin importar si se utilizó el método experimental o el método correlacional para formar los grupos, los cálculos son los mismos. También se pueden aplicar las fórmulas en los casos en que el

número de sujetos en cada grupo no sea igual (aunque ambos grupos deben tener números de sujetos aproximadamente iguales).

Cuando el diseño del experimento incluye una variable independiente con más de dos niveles (como en nuestro experimento del ejemplo, que incluyó tres grupos), el valor de *F* obtenido no nos indica si dos grupos específicos son significativamente distintos entre sí. Una forma de examinar las diferencias entre dos grupos en un estudio de este tipo, es utilizar la fórmula de SC_A para calcular la suma de cuadrados y el cuadrado medio de los dos grupos (los *gl* en este caso son 2-1). Cuando se hace esto, para calcular *F* debe utilizarse el CM_{ERROR} obtenido anteriormente, como la varianza del error. Se dispone de procedimientos más complejos para evaluar las diferencias entre dos grupos en este tipo de diseños, pero no son parte de los objetivos de este libro.

TABLA B.4 Tabla de resumen del análisis de varianza

Fuente de la varianza	Suma de cuadrados	*gl*	Cuadrados medios	*F*
A	SC_A	$a-1$	SC_A/gl_A	CM_A/CM_{ERROR}
Error	SC_{ERROR}	$N-a$	SC_{ERROR}/gl_{ERROR}	
Total	SC_{TOTAL}	$N-1$		
A	629.73	2	314.87	23.15
Error	163.20	12	13.60	
Total	792.93	14		

Análisis de varianza: dos variables independientes

En esta sección se describen los cálculos para el análisis de varianza con un diseño factorial que contiene dos variables independientes. Las fórmulas sirven para un diseño factorial *A* X *B,* con variables independientes que incluyan cualquier número de niveles. Las fórmulas funcionan únicamente con un diseño completamente de grupos independientes, con diferentes sujetos en cada grupo, y el número de sujetos en cada grupo debe ser el mismo. Sin embargo, una vez que usted comprenda este análisis, tendrá pocos problemas para entender el análisis de diseños más complejos con medidas repetidas o con distintas cantidades de sujetos. Con estas limitaciones en mente, consideremos los datos de un experimento hipotético.

El experimento utiliza un diseño factorial 2 x 2 VI x VP. La variable *A* es el tipo de enseñanza empleada en un curso, y la variable *B* es el nivel de inteligencia de los estudiantes, los cuales fueron clasificados con una inteligencia "baja" o "alta", con base en los puntajes de pruebas de inteligencia, y fueron asignados al azar a uno de dos tipos de clases. Una de las clases utiliza un método de enseñanza tradicional; la otra emplea un método de enseñanza individualizada, con

exámenes frecuentes de pequeñas cantidades de material, con monitores que ayudan a los estudiantes de forma individual y la estipulación de que los estudiantes deben manejar cada sección de material antes de pasar a la siguiente. La información que se les presenta en ambos tipos de clases es idéntica. Al final del curso, todos los estudiantes resuelven el mismo examen, que cubre todo el material presentado durante el curso. La calificación en el examen es la variable dependiente.

La tabla B.5 muestra datos ficticios de este experimento, con cinco sujetos en cada condición. Este diseño permite evaluar tres efectos: el efecto principal de A, el efecto principal de *B y* la interacción *A* x *B.* El efecto principal de *A* es si un tipo de enseñanza es superior a la otra; el efecto principal de *B* es si los estudiantes con un alto nivel de inteligencia obtienen distintas calificaciones en el examen que los estudiantes con un nivel bajo; la interacción *A* x *B* examina si el efecto de una de las variables independientes difiere de acuerdo con el nivel particular de la otra variable.

El cálculo del análisis de varianza se inicia con la obtención de la suma de cuadrados de las siguientes fuentes de varianza de los datos: SC_{T0TAI}, SC_A, SC_{qy} $SC_{A\ XB}$, *y* SC_{ERROR}- ^os Procedimientos para el cálculo son similares a los que se realizan para el análisis de varianza de una variable independiente. Los cálculos numéricos de los datos del ejemplo se presentan en la tabla B.6. Veamos ahora cada uno de estos cálculos.

SC_{TOTAL} La SC_{T0TAL} se calcula con la misma fórmula del análisis previo. La fórmula es

$$SC_{TOTAL} = \Sigma X^2 - \frac{G^2}{N}$$

donde ZZ^2 es la suma de los puntajes cuadrados de todos los sujetos en el experimento, G es el gran total de todos los puntajes y *N* es el número total de sujetos. Generalmente es más fácil calcular UC^2 y G en pasos más pequeños, al obtener los subtotales de forma separada para cada grupo en el diseño. Después se suman los subtotales. Este es el procedimiento que se siguió en las tablas B.5 yB.6.

SC_A La fórmula de la SC_A es

$$SC_A = \Sigma \frac{T_a^2}{n_a} - \frac{G^2}{N}$$

donde XT^2 es la suma de los totales cuadrados de los puntajes en cada grupo de la variable independiente *A, y* n_a es el número de sujetos en cada nivel de la variable independiente *A.* Al calcular SC_A sólo consideramos a los grupos de la variable independiente *A,* sin tomar en cuenta el nivel específico de *B.* En otras palabras,

TABLA B.5 Datos de un experimento hipotético sobre el efecto del tipo de enseñanza y el nivel de inteligencia, sobre la calificación en un examen: análisis de varianza

	Inteligencia (*B*)		
	Bajo (*B*1)	Alto (*B*2)	
Enseñanza tradicional (*A*1)	75 70 69 72 68	90 95 89 85 91	
	$T_{A1B1} = 354$ $\Sigma X^2_{A1B1} = 25094$ $n_{A1B1} = 5$ $\bar{X}_{A1B1} = 70.80$	$T_{A1B2} = 450$ $\Sigma X^2_{A1B2} = 40552$ $n_{A1B2} = 5$ $\bar{X}_{A1B2} = 90.00$	$T_{A1} = 804$ $n_{A1} = 10$ $\bar{X}_{A1} = 80.40$
Método individualizado (*A*2)	85 87 83 90 89	87 94 93 89 92	
	$T_{A2B1} = 434$ $\Sigma X^2_{A2B1} = 37704$ $n_{A2B1} = 5$ $\bar{X}_{A2B1} = 86.80$	$T_{A2B2} = 455$ $\Sigma X^2_{A2B2} = 41439$ $n_{A2B2} = 5$ $\bar{X}_{A2B2} = 91.00$	$T_{A2} = 889$ $n_{A2} = 10$ $\bar{X}_{A2} = 88.90$
	$T_{B1} = 788$ $n_{B1} = 10$ $\bar{X}_{B1} = 78.80$	$T_{B2} = 905$ $n_{B2} = 10$ $\bar{X}_{B2} = 90.50$	

los totales de cada grupo de la variable *A* se obtienen tomando en cuenta a todos los sujetos en ese nivel de A, sin importar en qué condición de *B* se encuentren los sujetos. G^2/N *se* calculó previamente para SC_{TOTAL}

SC_B La fórmula para SC_B es

$$SC_B = \frac{\Sigma T_b^2}{n_b} - \frac{G^2}{N}$$

SC_B se obtiene de la misma forma que SC_A. La única diferencia es que se calculan los totales de los grupos de la variable independiente *B.*

$SC_{A \times B}$ La fórmula para $SC_{A \times B}$, es

$$SC_{A \times B} = \frac{\Sigma T_{ab}^2}{n_{ab}} - \frac{G^2}{N} - SC_A - SC_B$$

La suma de cuadrados de la interacción *A x B se* obtiene calculando primero ZT^. Esto implica elevar al cuadrado el total de los puntajes de cada condición *ab* en el experimento. En nuestro ejemplo del experimento en la tabla B.5, existen cuatro condiciones; el cálculo de la interacción toma en cuenta a *todos* los grupos. Cada total por grupo se eleva al cuadrado y luego se obtiene la suma de los totales cuadrados. Esta suma se divide entre n_b, el número de sujetos en cada grupo. Las otras cantidades de la fórmula de SC_{AxB} ya se obtuvieron, por lo que su cálculo es relativamente directo.

TABLA B.6 Cálculos del análisis de varianza con dos variables independientes

$$SC_{TOTAL} = \Sigma X^2 - \frac{G^2}{N}$$
$$= (25094 + 40552 + 37704 + 41439) - \frac{(354+450+434+455)^2}{20}$$
$$= 144789 - 143312.45$$
$$= 1476.55$$

$$SC_A = \frac{\Sigma T_a^2}{n_a} - \frac{G^2}{N}$$
$$= \frac{(804)^2 + (889)^2}{10} - 143312.45$$
$$= 143673.70 - 143312.45$$
$$= 361.25$$

$$SC_B = \frac{\Sigma T_b^2}{n_b} - \frac{G^2}{N}$$
$$= \frac{(788)^2 + (905)^2}{10} - 143312.45$$
$$= 143996.90 - 143312.45$$
$$= 684.45$$

$$SC_{A \times B} = \frac{\Sigma T_{ab}^2}{n_{ab}} - \frac{G^2}{N} - SC_A - SC_B$$
$$= \frac{(354)^2 + (450)^2 + (434)^2 + (455)^2}{5} - 143312.45 - 361.25 - 684.45$$
$$= 144639.40 - 143312.45 - 361.25 - 684.45$$
$$= 281.25$$

$$SC_{ERROR} = \Sigma X^2 - \frac{\Sigma T_{ab}^2}{n_{ab}}$$
$$= 144789 - 144639.40$$
$$= 149.60$$

SE_{ERROR} Los valores necesarios para SC_{ERR0R} ya fueron calculados. La fórmula es

$$SC_{ERROR} = \Sigma X^2 - \frac{\Sigma T_{ab}^2}{n_{ab}}$$

Estos valores se calcularon antes, por lo que únicamente necesitamos realizar la resta correspondiente para completar el cálculo de SC_{ERR0R}.

En este punto, tal vez usted quiera practicar el cálculo de las sumas de cuadrados con los datos de la tabla B.5. Para verificar los resultados, asegúrese de que $SC_{TOTAL} = SC_A + SC_B + SC_{AxB} + SC_{ERR0R}$.

Después de obtener las sumas de cuadrados, el siguiente paso es encontrar el cuadrado medio de cada una de las fuentes de la varianza. La forma más sencilla de hacerlo es mediante el uso de una tabla de resumen del análisis de varianza, como la tabla B.7.

Cuadrados medios El cuadrado medio de cada una de las fuentes de la varianza es la suma de cuadrados, dividida entre los grados de libertad. Las fórmulas de los grados de libertad y de los cuadrados medios se muestran en la parte superior de la tabla B.7 y los valores calculados se muestran en la parte inferior.

Obtención del valor *F* El valor *F* de cada una de las tres fuentes de varianza sistemática (efectos principales para A y *B y* la interacción) se obtienen al dividir el cuadrado medio apropiado entre el CM_{ERR0R}. Ahora tenemos tres valores de *F y* podemos evaluar la significación de los efectos principales y de la interacción.

Significación de *F* Para determinar si un valor obtenido de *F* es significativo, necesitamos encontrar el valor crítico de *F* en la tabla C.4 del apéndice C. Para

TABLA B.7 Tabla de resumen del análisis de varianza: dos variables independientes

Fuente de varianza	Suma de cuadrados	*gl*	Cuadrados medios	*F*
A	SC_A	$a - 1$	SC_A/gl_A	CM_A/CM_{ERROR}
B	SC_B	$B - 1$	SC_B/gl_B	CM_B/CM_{ERROR}
$A \times B$	$SC_{A \times B}$	$(a - 1)(b - 1)$	$SC_{A \times B}/gl_{A \times B}$	$CM_{A \times B}/CM_{ERROR}$
Error	SC_{ERROR}	$N - ab$	SC_{ERROR}/gl_{ERROR}	
Total	SC_{TOTAL}			
A	361.25	1	361.25	38.64
B	684.45	1	684.45	73.20
$A \times B$	281.25	1	281.25	30.08
Error	149.60	16	9.35	
Total	1476.55	19		

todos los valores de *F* en la tabla del resumen del análisis de varianza, los grados de libertad son 1 y 16. Suponga que se eligió un nivel de significación de 0.01 para el rechazo de la hipótesis nula. El valor crítico de *F,* al nivel 0.01, para 1 y 16 grados de libertad, es 8.53. Si la *F* obtenida es mayor que 8.53, podemos decir que los resultados son significativos a un nivel de 0.01. Si se remite a los valores de *F* obtenidos en la tabla B.7, observará que los efectos principales y la interacción son significativos. Interprete usted las medias de los efectos principales y grafique la interacción. Si no recuerda cómo hacer esto, repase el material del capítulo 10.

Análisis de varianza.- medidas repetidas y sujetos apareados

Los cálculos para el análisis de varianza presentados hasta aquí, se han limitado a diseños de grupos independientes. Esta sección introduce el análisis de varianza para diseños de medidas repetidas o para una asignación aleatoria apareada con una variable independiente.

En la tabla B.8 se incluyen datos ficticios de un experimento hipotético con medidas repetidas. El experimento examina el efecto del atractivo físico de un candidato a un empleo, sobre los juicios que se hacen sobre su capacidad. La variable independiente es el atractivo físico del candidato; la variable dependiente son los juicios sobre su capacidad, en una escala de 10 puntos. Los participantes en el experimento observaron dos filmaciones de dos mujeres realizando una tarea de habilidades mecánicas que implicaba unir ciertas piezas. Ambas mujeres se desempeñaron igualmente bien, pero una era físicamente atractiva y la otra no. El orden de la presentación de las dos filmaciones se contrabalanceó para controlar los efectos de orden.

La diferencia principal entre el análisis de varianza de medidas repetidas y el análisis de grupos independientes, descrito anteriormente, consiste en que el efecto de las diferencias entre los sujetos se convierte en una fuente de varianza. Hay cuatro fuentes de varianza en el análisis de varianza de medidas repetidas, y por lo tanto, se calculan cuatro sumas de cuadrados:

$$SC_{TOTAL} = \Sigma X^2 - \frac{G^2}{N}$$

$$SC_A = \frac{\Sigma T_a^2}{n_a} - \frac{G^2}{N}$$

$$SC_{SUJETOS} = \frac{\Sigma T_s^2}{n_s} - \frac{G^2}{N}$$

$$SC_{ERROR} = SC_{TOTAL} - SC_A - SC_{SUJETOS}$$

Los cálculos para estas sumas de cuadrados se incluyen en la parte inferior de la tabla B.8. Usted debe estar familiarizado con las cifras de la fórmula. La única

TABLA B.8 Datos de un experimento hipotético sobre el atractivo físico y los juicios de las capacidades: análisis de varianza de medidas repetidas

	Condición (*A*)			
Sujetos (o pares de sujetos)	Candidata poco atractiva (*A*1)	Candidata atractiva (*A*2)	T_s	T_s^2
#1	6	8	14	196
#2	5	6	11	121
#3	5	9	14	196
#4	7	6	13	169
#5	4	6	10	100
#6	3	5	8	64
#7	5	5	10	100
#8	4	7	11	121
	$T_{A1} = 39$	$T_{A2} = 52$		$\Sigma T_s^2 = 1067$
	$\Sigma X_{A1}^2 = 201$	$\Sigma X_{A2}^2 = 352$		
	$n_{A1} = 8$	$n_{A2} = 8$		
	$\overline{X}_{A1} = 4.88$	$\overline{X}_{A2} = 6.50$		

$$SC_{TOTAL} = \Sigma X^2 - \frac{G^2}{N} = (201 + 352) - \frac{(39+52)^2}{16}$$
$$= 553 - 517.56$$
$$= 35.44$$

$$SC_A = \frac{\Sigma T_a^2}{n_a} - \frac{G^2}{N} = \frac{(39)^2 + (52)^2}{8} - 517.56$$
$$= 528.13 - 517.56$$
$$= 10.57$$

$$SC_{SUJETOS} = \frac{\Sigma T_s^2}{n_s} - \frac{G^2}{N} = \frac{1067}{2} - 517.56$$
$$= 533.50 - 517.56$$
$$= 15.94$$

$$SC_{ERROR} - SC_{TOTAL} - SC_A - SC_{SUJETOS} = 35.44 - 10.57 - 15.94$$
$$= 8.93$$

cifra nueva implica el cálculo de $SC_{SUfET0S}$. El símbolo T^2_s se refiere al cuadrado del puntaje total de cada sujeto, es decir, el total cuadrado de los puntajes que cada sujeto proporciona al ser medido en los distintos grupos del experimento. El valor $\sum T^2$ se refiere a la sumatoria de estos totales cuadrados de todos los

sujetos. El cálculo de $SC_{SUJETOS}$ se completa al dividir $\sum T^2$ entre n_s y después restar el resultado de G^2/N. El término n se refiere al número de puntajes que proporciona cada sujeto. Puesto que nuestro experimento hipotético incluye dos grupos, $n_s = 2$, el total de cada sujetos se basa en dos puntajes.

En la tabla B.9 se muestra el resumen del análisis de varianza. Los procedimientos para el cálculo de los cuadrados medios y de F son similares a los ejemplos anteriores. Observe que no se obtuvo el cuadrado medio ni la F de la fuente de varianza de los sujetos; generalmente no hay una razón para preocuparse o conocer si los sujetos difirieron de manera significativa entre sí. La posibilidad de calcular esta fuente de varianza tiene la ventaja de reducir la varianza del error; en un diseño de grupos independientes, las diferencias entre los sujetos conforman la varianza del error. Dado que sólo hay un puntaje por sujeto en el diseño de grupos independientes, es imposible estimar la influencia de las diferencias entre los sujetos.

Puede utilizar la tabla del resumen y la tabla de los valores críticos de F para determinar si la diferencia entre los dos grupos es significativa. Los procedimientos son idénticos a los que explicamos con anterioridad.

Análisis de varianza: conclusión

El análisis de varianza es una prueba muy útil que puede aplicarse a cualquier tipo de diseño factorial, incluyendo a los que emplean grupos independientes y medidas repetidas en el mismo diseño. El método para el cálculo del análisis de varianza es casi el mismo, sin importar la complejidad del diseño. Una sección dedicada al análisis de varianza tan breve como ésta no puede cubrir todos los aspectos de esta técnica estadística general. Sin embargo, ahora usted debe tener los fundamentos para realizar este análisis y para comprender las discusiones más detalladas en libros de estadística avanzada.

TABLA B.9 Tabla de resumen del análisis de varianza: diseño de medidas repetidas

Fuente de varianza	Suma de cuadrados	*gl*	Cuadrados medios	*F*
A	SC_A	$a - 1$	SC_A/gl_A	CM_A/CM_{ERROR}
Sujetos	$SC_{SUJETOS}$	$s - 1$	—	
Error	SC_{ERROR}	$(a - 1)(s - 1)$	SC_{ERROR}/gl_{ERROR}	
Total	SC_{TOTAL}	$N - 1$		
A	10.57	1	10.57	8.26
Sujetos	15.94	7	—	
Error	8.93	7	1.28	
Total	35.44	15		

CORRELACIÓN Y MAGNITUD DEL EFECTO

Finalmente, examinaremos los cálculos para las medidas de correlación y de la magnitud del efecto; éstas son indicadores de la fuerza de la asociación entre variables y son muy importantes debido a que proporcionan una métrica común que puede utilizarse para todo tipo de estudios. Estos valores varían desde 0.00, que indica la ausencia de relación, hasta 1.00; las correlaciones superiores a 0.50 se consideran indicadoras de relaciones muy fuertes. En muchas investigaciones se esperan correlaciones de entre 0.25 y 0.50. Correlaciones de entre 0.10 y 0.25 son más débiles, aunque pueden ser estadísticamente significativas en muestras grandes; también pueden ser importantes por razones teóricas, e incluso prácticas.

Magnitud del efecto del estadístico chi cuadrada

Anteriormente describimos la prueba chi cuadrada (X^2). Además de determinar si existe una relación significativa, se puede buscar un indicador de la magnitud del efecto que señale la fuerza de la relación entre las variables. Este se obtiene por medio del coeficiente Phi, que se simboliza Φ. Phi se calcula después de obtener el valor de chi cuadrada. La fórmula es:

$$\phi = \sqrt{\frac{x^2}{N}}$$

Así, el valor de Φ para el estudio del género y la mano dominante, analizado previamente (véase la tabla B.2), es:

$$\phi = \sqrt{\frac{18.00}{100}} = \sqrt{0.18} = 0.42$$

Puesto que ya se determinó la significación del valor de chi cuadrada, no se requiere de más pruebas de significación. Nota: con tablas más grandes se utiliza un estadístico relacionado, llamado V de Cramer.

Magnitud del efecto para el estadístico F

Después de realizar el análisis de varianza y evaluar la significación del estadístico i[7], es necesario examinar la magnitud del efecto. *Eta* es un tipo de coeficiente de correlación que puede calcularse con facilidad. La fórmula es

$$\text{eta} = \sqrt{\frac{\text{varianza entre grupos (sistemática)}}{\text{varianza total}}}$$

En el experimento de la distancia interpersonal y la revelación de información, descrito antes, la SC_A fue 629.73 y la SC_{TOTAL} fue 792.93. Entonces, el valor de eta sería

$$\text{eta} = \sqrt{\frac{629.73}{792.93}}$$

$$= 0.89$$

Esta es una correlación muy alta, que refleja el hecho de que los datos fueron fabricados para facilitar los cálculos.

Coeficiente de correlación producto-momento de Pearson

El coeficiente de correlación r producto-momento de Pearson se usa para determinar la fuerza de la relación entre dos variables medidas en una escala de intervalo o de razón.

Ejemplo Suponga que desea conocer si los viajes están relacionados con los conocimientos de geografía. En su estudio, aplica un cuestionario con 15 reactivos, acerca de la geografía de México, y también pregunta cuántos Estados han visitado los participantes. Después de obtener los pares de puntajes de cada participante, se puede calcular una *r* de Pearson para medir la fuerza de la relación ente los viajes y los conocimientos de geografía.

La tabla B.10 incluye datos ficticios de un estudio de este tipo, junto con los cálculos de r. La fórmula para obtener r es

$$r = \frac{N\Sigma XY - \Sigma X / \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

donde *X* se refiere al puntaje de un sujeto en la variable X, y Fes el puntaje de un sujeto en la variable *Y.* En la tabla B.10 los puntajes del número de viajes constituyen la variable *X y* los puntajes del conocimiento de geografía la *variable Y.* En la fórmula, *N* es el número de pares de observaciones (es decir, el número de participantes medidos en ambas variables).

El cálculo de *r* requiere ciertas operaciones aritméticas con los puntajes *de X y Y.* $\sum X$ es simplemente la suma de los puntajes de la variable *X.* $\sum X^2$ es la suma del cuadrado de los puntajes de *X* (primero se eleva al cuadrado cada puntaje y después se obtiene la suma de los puntajes cuadrados). El valor de $(\sum X)^2$ es el cuadrado de la suma de los puntajes: primero se calcula el total de los puntajes $X(\sum X)$ (y luego se eleva al cuadrado este total. Es importante no confundir las dos cantidades, $\sum X^2$ *y* $(\sum X)^2$. Se hacen los mismos cálculos con los puntajes *Y,* para obtener, $\sum Y$, $\sum Y^2$ y $(\sum Y)^2$. Para calcular $\sum XY$, se multiplica cada puntaje *X* por el puntaje *Y* de cada participante; luego se suman estos valores de todos los

TABLA B.1O Datos de un estudio hipotético sobre los viajes y los conocimientos en geografía: *r* de Pearson

Número de identificación del sujeto	Puntuación de viaje (X)	Puntuación de conocimientos (Y)	XY
01	4	10	40
02	6	15	90
03	7	8	56
04	8	9	72
05	8	7	56
06	12	10	120
07	14	15	210
08	15	13	195
09	15	15	225
10	17	14	238
	$\Sigma X = 106$	$\Sigma Y = 116$	$\Sigma XY = 1302$
	$\Sigma X^2 = 1308$	$\Sigma Y^2 = 1434$	
	$(\Sigma X)^2 = 11236$	$(\Sigma Y)^2 = 13456$	

Cálculo:

$$r = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{10(1302) - (106)(116)}{\sqrt{10(1308) - 11236}\sqrt{10(1434) - 13456}}$$

$$= \frac{13020 - 12296}{\sqrt{13080 - 11236}\sqrt{14340 - 13456}}$$

$$= \frac{724}{\sqrt{1844}\sqrt{884}}$$

$$= \frac{724}{1276.61}$$

$$= 0.567$$

sujetos. Cuando se hayan obtenido todos estos resultados, se calcula *r* por medio de la fórmula anterior.

En este punto, debe examinar cuidadosamente los cálculos que se muestran en la tabla B.10, para que se familiarice con los procedimientos del cálculo de r. Después puede intentar obtener r de otro conjunto de datos, como el asiento y la calificación del examen de la tabla 12.2.

Significación de r Para probar la hipótesis nula que establece que el coeficiente de correlación poblacional es 0.00, se consulta la tabla de valores críticos de r. La tabla C.5, del apéndice C, muestra los valores críticos de r para los niveles 0.10, 0.05 y 0.01 de significación. Para encontrar el valor crítico, primero necesita determinar los grados de libertad. Los *gl* para la prueba de significación de *r* son *N* - 2. En nuestro ejemplo de los viajes y los conocimientos, el número de pares de observaciones es 10, por lo que *gl* = 8. Para 8 grados de libertad, el valor crítico de r, a un nivel de significación de 0.05, es 6.32 (+ o -). La r obtenida debe ser mayor que el valor crítico para ser significativa. Dado que nuestra *r* obtenida (de la tabla B.10) de 0.567 es menor que el valor crítico, no se rechaza la hipótesis nula.

Observe que no se rechaza la hipótesis nula en este caso, aun cuando la magnitud de *r* es bastante grande. Recuerde la explicación acerca de los resultados no significativos del capítulo 13. Es probable que obtuviera una correlación significativa si eligiera una muestra mayor o medidas más sensibles y confiables de las variables.

Apéndice C
Tablas estadísticas

TABLA DE NÚMEROS ALEATORIOS

Los número aleatorios entre 0 y 99, que aparecen en la Tabla C.l, pueden utilizarse para asignar a los participantes de forma aleatoria a las condiciones en un experimento o para seleccionar aleatoriamente a los individuos de una población. Para obtener una serie de números aleatorios, elija un punto arbitrario en la tabla y lea la secuencia hacia un lado o hacia abajo.

Para una asignación aleatoria, primero ordene a los participantes de alguna forma. Por ejemplo, suponga que en su experimento participarán 15 individuos. Podría simplemente ordenarlos del primero al último, como se muestra en la siguiente tabla. Ahora puede emplear la tabla de números aleatorios para asignarlos a las condiciones del experimento. Suponga que necesita asignar a estos 15 individuos a tres grupos. Utilice la tabla de números aleatorios y asigne un número a cada uno de los 15 participantes. (Si un número se duplica, ignórelo y utilice el siguiente número en la secuencia). En el ejemplo que se incluye aquí, se inició en la parte superior izquierda de la tabla y se leyó hacia abajo. Ahora asigne a los participantes a los tres grupos: los cinco que tengan los números aleatorios más bajos se incluyen en el Grupo 1, los siguientes cinco en el Grupo 2 y los cinco individuos con los números más altos se asignan al Grupo 3. Estos procedimientos generales pueden aplicarse a cualquier cantidad de grupos en un experimento.

Para utilizar la tabla de números aleatorios en un muestreo, primero haga una lista de todos los miembros de su población. Use la tabla de números aleatorios y asigne un número a cada miembro de la población. Determine el tamaño muestral (N), y entonces su muestra estará compuesta de los primeros *N* individuos. Por ejemplo, si desea tomar una muestra aleatoria de 25 profesores de su escuela, utilice la tabla de números aleatorios para asignarle un número a cada uno de ellos. Los 15 profesores con los números más bajos se seleccionarán para la muestra.

Ahora existen alternativas por computadora a las tablas de números aleatorios, como la tabla C.l. Programas de hojas de cálculo como Excel, incluyen una función para producir números aleatorios. También existe un útil sitio de Internet para asignación y selección aleatorias en http://www.randomizer.org

Orden de los participantes	Número aleatorio	Grupo asignado
1	56	2
2	*57*	2
3	51	2
4	10	1
5	69	2
6	9	1
7	75	3
8	5	1
9	78	3
10	90	3
11	79	3
12	4	1
13	79	3
14	48	1
15	57	2

TABLA C.l Tabla de números aleatorios

56	54	77	81	67	11	27	56	76	67	16	71	93	68	33	31	90	67	52	32	81	64	3
57	63	49	93	11	26	57	88	94	65	80	23	82	16	58	86	90	70	54	71	13	50	60
51	1	35	56	36	28	18	97	49	47	80	43	56	73	22	69	58	37	55	52	56	69	34
10	34	73	74	9	24	14	31	97	62	15	80	51	49	64	48	61	9	14	29	81	65	16
69	59	39	53	82	97	70	98	90	22	12	22	13	87	10	69	61	11	27	60	50	42	96
9	94	34	81	9	79	41	30	35	27	54	82	49	95	79	3	26	2	10	12	93	15	12
75	9	76	33	71	87	5	39	13	98	68	43	32	46	52	75	27	71	66	62	37	70	14
5	54	84	29	60	75	35	6	31	82	39	31	98	45	97	73	56	3	39	48	36	49	4
78	23	95	76	61	61	12	52	68	57	72	64	65	12	88	42	8	29	58	56	10	32	32
90	18	61	31	11	45	44	6	15	90	74	69	72	38	94	35	75	27	75	76	25	91	33
79	70	44	70	2	61	6	4	49	96	17	37	3	78	63	82	57	50	24	55	43	53	84
4	53	53	85	61	99	99	0	34	17	24	65	59	66	75	80	16	95	96	25	61	36	2
79	50	9	86	37	60	76	80	13	80	78	16	46	87	56	6	71	17	96	58	63	80	43
48	13	20	30	35	34	99	51	32	90	6	8	56	61	12	94	86	6	51	89	53	26	13
57	66	6	29	10	3	33	73	0	69	19	95	35	2	52	88	31	49	40	53	90	46	10
82	47	37	12	59	51	55	24	35	83	27	34	53	50	87	15	2	20	31	65	15	91	7
64	38	48	70	7	46	86	42	24	34	68	58	84	54	41	13	23	38	69	29	22	5	27
98	19	18	8	57	56	76	69	92	64	69	11	79	93	98	3	70	76	6	72	70	83	28
4	53	16	78	53	54	31	33	2	64	29	24	45	49	30	82	24	25	81	25	22	43	3
51	12	24	741	1	35	67	55	45	43	88	54	52	90	10	73	92	8	25	0	59	34	75
31	33	22	54	64	88	52	70	43	44	84	43	42	6	57	34	25	34	90	95	54	25	18
13	73	99	45	24	5	54	90	55	99	24	67	50	58	52	33	36	84	9	92	79	77	5
73	12	17	38	11	41	25	86	12	85	24	18	3	27	80	54	59	92	51	71	34	58	65
10	80	53	35	0	10	52	55	19	94	20	43	63	48	44	2	70	26	94	28	92	38	86
12	5	47	41	45	77	99	91	87	9	34	67	63	61	22	63	37	17	87	89	80	27	24
34	64	18	70	39	71	85	67	11	70	55	33	21	72	84	92	43	99	6	14	77	36	54
40	7	10	35	99	37	94	70	17	75	34	82	27	47	62	2	61	69	99	57	50	34	36
24	46	26	12	98	54	6	83	73	11	69	11	30	90	69	15	5	90	2	40	90	78	64
85	73	53	28	25	32	79	24	29	54	77	69	38	45	9	3	14	5	42	26	57	51	89
10	34	23	44	5	98	79	55	92	99	49	53	12	24	42	27	83	69	77	92	90	13	28
69	17	91	38	2	46	98	5	97	22	62	19	48	96	38	11	5	15	30	79	33	64	7
18	71	29	18	3	17	56	39	69	19	40	69	68	64	98	7	32	33	52	33	95	7	57
32	28	12	9	99	92	52	98	56	5	12	71	42	23	3	64	4	5	33	62	5	34	46
61	57	16	77	65	62	39	89	68	82	51	42	91	73	53	25	59	1	93	57	28	4	17
10	66	15	18	62	62	79	88	79	84	47	21	86	0	6	21	57	66	12	55	72	31	6
83	32	65	27	30	50	32	60	81	99	86	83	90	2	70	82	55	24	1	79	12	64	32
82	69	70	0	64	77	55	41	71	42	5	40	76	10	74	62	24	36	8	85	0	24	25
94	15	11	46	59	85	81	54	89	93	70	20	4	79	9	18	5	48	77	52	84	32	68
57	83	28	41	10	6	11	22	48	33	34	35	51	77	38	7	26	34	72	49	52	59	47
76	58	72	63	41	58	67	40	79	28	36	63	27	79	47	97	61	12	84	97	7	97	41

4	52	15	69	32	65	40	39	11	63	30	53	29	21	70	67	57	85	92	95	64	72	61
86	4	49	39	55	71	47	73	78	42	48	32	68	13	65	43	54	44	61	32	16	21	10
15	98	53	34	70	37	65	88	0	58	3	65	57	45	5	43	93	18	19	31	4	86	69
10	21	90	40	38	58	1	28	18	98	40	37	1	19	76	62	54	29	6	27	96	61	34
20	0	53	71	35	49	81	60	51	87	25	13	16	77	31	43	60	49	27	66	1	56	9
28	48	61	31	66	66	26	62	90	11	86	34	55	95	26	12	39	20	51	85	77	18	81
63	34	49	96	18	31	59	61	26	18	34	86	31	97	23	77	44	25	24	31	79	94	8
63	77	79	43	44	26	3	13	21	56	31	71	71	57	23	0	15	57	57	46	53	38	59
46	74	49	18	77	16	74	10	55	23	2	50	82	63	17	76	63	99	40	26	35	89	42
58	28	44	4	39	69	71	59	68	28	18	18	6	4	99	75	64	64	87	59	11	37	42

46	76	17	53	46	94	82	43	71	41	30	13	51	48	28	52	97	13	19	70	58	37	55
45	98	45	44	70	83	73	99	86	85	4	86	45	29	80	94	78	79	32	25	76	5	70
27	70	86	16	32	59	99	71	31	22	62	44	39	80	52	27	30	82	73	23	51	36	79
38	32	43	34	10	3	36	24	91	24	26	62	19	70	31	53	69	1	73	52	6	8	21
37	79	13	76	64	55	48	79	84	81	19	48	71	72	47	85	75	7	6	41	74	51	59
46	98	85	52	65	48	9	36	54	50	55	1	92	7	2	50	4	35	35	39	66	71	31
66	12	24	53	23	83	97	13	61	93	44	12	15	46	50	56	15	12	20	84	23	89	64
60	6	52	97	15	78	31	36	41	89	52	71	63	1	23	16	60	78	86	47	15	3	82
99	73	76	80	73	84	6	51	96	76	50	25	14	61	17	46	1	74	43	71	72	99	73
88	58	26	18	5	60	0	41	6	58	65	43	0	57	91	2	9	30	44	78	94	95	72

92	94	3	46	42	12	52	26	95	48	71	74	98	50	12	69	62	42	20	50	1	1	70
95	64	66	18	69	33	82	71	32	56	42	63	35	56	96	18	3	14	80	25	80	41	84
68	19	20	17	34	27	27	51	97	71	35	99	2	37	25	81	52	39	91	4	36	28	31
57	96	79	5	89	24	90	32	69	37	2	16	38	89	22	25	9	87	21	82	87	24	90
60	45	57	82	98	25	41	41	35	8	84	62	55	90	59	15	35	63	14	87	69	86	3
40	8	48	28	46	86	71	71	56	79	61	52	33	38	74	90	87	82	17	11	72	31	79
90	32	23	16	91	48	32	76	77	68	97	84	72	31	23	44	37	90	97	17	11	76	88
96	61	45	64	81	15	33	99	55	92	69	3	46	80	64	48	32	76	47	39	27	2	31
38	66	94	89	57	57	31	84	64	11	51	55	55	4	11	24	80	94	76	81	42	65	90
72	69	51	48	31	9	46	32	13	23	95	93	34	82	78	0	97	75	65	30	21	97	24

76	15	55	72	55	21	5	90	12	20	85	80	57	80	17	93	86	78	30	18	30	48	77
16	17	89	69	42	75	33	11	22	7	12	43	38	58	89	41	58	4	34	73	83	20	10
30	67	23	66	19	43	84	14	6	26	19	22	0	68	47	82	77	5	7	22	76	37	6
7	84	19	23	54	13	36	62	49	14	91	61	46	20	2	15	44	22	55	48	87	18	5
69	74	59	7	66	96	42	21	16	87	44	47	87	99	60	63	83	51	14	93	33	41	76
92	80	5	3	81	7	10	56	50	2	22	90	93	82	39	42	52	74	52	79	33	14	20
43	46	11	66	75	70	26	19	39	24	68	16	77	81	30	90	38	88	34	90	93	96	98
21	36	36	12	35	27	95	98	2	32	1	43	20	77	24	26	96	0	57	39	70	85	92
26	41	22	42	17	8	97	67	74	13	87	63	5	81	8	18	23	39	94	28	25	34	14
98	3	56	19	5	83	82	23	98	7	0	59	73	84	82	70	12	8	51	73	20	5	86

(continúa)

TABLA C.1 Tabla de números aleatorios *(continuación)*

28	33	51	32	6	50	6	86	61	4	68	0	15	80	16	15	82	91	61	22	47	2	23
30	12	3	27	98	73	17	61	91	82	56	24	9	34	63	44	89	43	41	73	80	2	91
10	34	45	83	27	17	21	70	66	73	14	3	84	92	53	56	0	11	74	19	31	38	47
55	72	22	83	89	32	48	41	48	88	90	84	1	11	35	8	93	43	45	13	78	47	31
39	61	21	65	72	49	63	5	65	64	12	25	69	8	32	61	86	15	16	87	35	89	72
49	74	40	83	29	55	79	69	1	79	73	9	73	88	16	10	3	34	25	54	8	27	22
11	65	53	34	53	78	51	55	0	85	91	68	17	46	4	72	90	8	56	88	24	85	96
10	63	78	95	46	99	9	37	42	15	53	59	80	26	15	74	82	60	30	10	74	74	70
49	86	40	80	73	20	64	72	55	12	17	80	45	95	8	40	70	87	80	1	9	30	19
60	4	79	64	53	28	9	10	91	24	44	27	5	55	15	96	17	11	1	26	87	20	38

16	99	24	90	2	63	59	39	21	4	16	26	80	12	12	83	28	7	10	93	65	85	54
65	14	4	40	99	87	73	41	65	21	81	94	22	23	52	0	25	73	10	74	92	87	24
21	15	19	37	3	52	87	27	30	16	22	82	20	46	16	40	27	90	89	59	93	6	28
56	82	81	89	30	12	51	48	37	16	94	31	97	36	23	93	82	16	89	29	96	80	99
1	60	99	69	11	20	9	5	84	82	10	20	87	61	87	48	95	5	83	28	24	6	87
60	11	63	82	24	30	29	95	38	30	70	66	78	13	38	41	16	60	63	39	44	70	8
78	5	70	38	3	49	85	15	42	47	12	20	1	6	7	87	61	73	91	19	37	11	73
25	16	0	86	71	60	16	27	80	27	26	89	94	46	22	51	62	29	83	21	90	31	71
66	33	96	63	95	69	53	61	40	37	73	42	16	62	13	0	20	85	4	71	25	1	43
51	20	42	96	94	96	70	53	71	48	26	84	94	37	62	31	19	84	28	88	61	40	58

73	33	16	26	56	63	91	2	13	81	60	13	88	31	86	63	14	16	24	35	87	35	37
21	91	74	11	28	24	62	61	44	65	90	27	14	92	10	28	22	50	22	68	73	79	3
9	61	60	69	97	11	85	59	0	6	59	16	12	7	85	2	16	20	39	59	33	81	14
16	62	92	65	82	61	26	8	36	86	9	73	68	79	60	10	32	48	59	70	83	95	5
25	45	23	46	19	70	36	7	75	26	41	14	36	32	25	87	34	30	74	57	39	61	16
85	83	20	36	4	20	69	12	59	32	53	33	33	68	38	37	18	4	38	63	16	73	85
31	63	91	65	30	40	27	30	15	65	31	43	58	99	30	53	97	91	95	2	89	53	0
54	94	59	61	69	98	22	50	74	63	3	39	46	22	35	40	4	11	61	14	21	57	2
82	86	93	17	2	30	12	23	28	24	34	41	74	53	35	88	0	19	71	6	59	38	0
32	93	35	64	4	86	60	12	80	30	45	75	18	41	11	47	99	30	55	72	53	99	92

82	79	45	35	16	95	16	44	36	95	36	88	13	55	26	81	3	55	42	65	65	12	93
85	13	56	62	53	14	35	18	87	26	27	17	72	79	83	81	89	64	89	33	86	52	16
3	21	83	18	41	64	54	36	89	49	46	11	6	87	97	11	12	36	59	75	81	59	20
15	98	56	37	29	75	56	72	44	28	8	11	45	74	44	51	75	90	63	55	20	97	47
15	0	33	3	96	88	88	88	97	62	36	95	92	94	4	63	61	11	21	42	62	20	21
39	51	58	87	73	15	47	34	81	60	99	40	18	50	86	80	88	99	11	79	93	19	27
22	42	18	32	49	40	72	4	46	56	2	60	59	91	47	22	27	11	35	66	72	88	11
87	73	0	50	32	20	7	46	65	85	43	73	97	46	73	65	0	8	38	34	11	0	81
85	99	68	83	61	9	29	40	78	71	93	93	61	56	36	50	16	65	2	83	1	41	27
34	10	45	58	47	11	9	20	37	33	48	82	9	73	6	2	19	51	99	44	83	49	82

24	30	59	32	34	58	6	54	13	2	86	18	65	26	61	61	38	38	72	60	55	92	88
76	92	62	70	77	89	0	75	5	2	44	84	28	51	64	40	82	60	39	4	88	49	45
41	27	22	49	79	21	20	58	10	25	74	73	62	81	99	99	61	21	18	66	96	81	43
83	71	66	90	13	11	39	47	79	32	23	53	42	17	45	82	3	74	22	47	5	46	76
68	49	85	27	2	8	25	5	9	93	61	68	2	64	50	82	8	43	48	1	21	52	13
86	26	41	11	48	77	93	15	54	99	95	71	81	91	8	26	67	73	52	5	2	66	24
25	58	61	79	-7	48	8	14	8	78	93	56	70	45	92	58	31	10	92	0	12	47	70
73	35	64	61	48	69	14	40	24	91	84	56	83	36	50	59	0	72	58	44	57	96	34
41	12	91	81	45	33	73	79	21	26	58	0	20	2	37	61	17	19	87	28	16	70	88
81	80	97	64	80	19	31	11	85	23	76	28	72	71	3	43	50	39	64	89	41	65	54

41	40	11	55	45	99	71	7	55	75	31	73	41	0	18	36	31	45	88	67	40	51	83
50	15	71	99	43	88	59	6	94	75	95	58	7	30	80	99	12	91	57	53	90	26	12
77	54	33	43	1	15	86	91	58	84	97	65	75	46	25	39	80	59	95	75	60	27	51
79	66	68	73	40	96	6	15	21	97	83	25	30	69	36	19	91	40	86	62	28	92	56
21	3	62	2	69	74	50	6	83	92	93	86	65	80	74	25	51	93	46	74	94	25	14
82	62	93	95	44	83	31	29	78	41	24	74	41	10	29	14	49	43	83	46	83	76	46
58	72	90	3	89	15	64	72	45	15	73	14	14	47	3	29	1	19	75	70	3	16	73
81	57	48	9	34	27	23	81	44	48	26	7	19	74	43	71	46	11	98	5	14	64	78
65	65	86	84	54	33	80	15	77	34	85	97	42	98	72	31	88	42	23	47	78	54	45
90	75	83	5	91	6	95	23	89	44	5	96	61	7	35	88	71	97	59	72	64	73	85

TABLA C.2 Valores críticos de chi cuadrada

Grados de libertad	Nivel de probabilidad		
	.10	.05	.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566

Fuente: Tabla adaptada de Fisher y Yates, *Statistical Tables For Biológica!, Agñcuiíural and Medical Research* (1974. 6a. ed.). Londori: Longman. Reproducida con permiso.

TABLA C.3 Valores críticos de t

	Nivel de significación*			
	.05	.025	.01	.005
gl	.10	.05	.02	.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.875	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576

*Use el nivel de significación de arriba cuando haya predicho una diferencia direccional específica (una prueba de una cola; *e.* G., el Grupo 1 será mayor que el Grupo 2). Utilice el nivel de significación de abajo cuando sólo haya predicho que el Grupo 1 diferirá del Grupo *2,* sin especificar la dirección de la diferencia (una prueba de dos colas).

TABLA C.4 Valores críticos de *F*

g/del denominador (del error)	a	*gl* del numerador (sistemática)											
		1	2	3	4	5	6	7	8	9	10	11	12
1	.10	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
	.05	161	200	216	225	230	234	237	239	241	242	243	244
2	.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
	.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
	.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
	.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
	.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1
4	.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
	.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4
5	.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	4.34	3.32	3.30	3.28	3.27
	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
	.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89
6	.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
	.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	.10	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
	.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47
8	.10	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.10	3.44	3.39	3.35	3.31	3.28
	.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
9	.10	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
	.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11
10	.10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
	.05	4.96	4.10	3.71	3.48	3.33	3.22	4.14	3.07	3.02	2.98	2.94	2.91
	.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
11	.10	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40

TABLA C.4 Valores críticos dé *F (continuación)*

gl del denominador (del error)	a	*gl* del numerador (sistemática)											
		1	2	3	4	5	6	7	8	9	10	11	12
12	.10	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.;69
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16
13	.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	.10	3.10	2.76	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.08	2.05
	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	.10	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
17	.10	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46
18	.10	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.96	1.93
	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
19	.10	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.94	1.91
	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	.10	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.92	1.89
	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23
22	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03

{continúa)

TABLA C.4 Valores críticos de *F (continuación)*

gl del denominador (del error)	a	*gl* del numerador (sistemática)											
		1	2	3	4	5	6	7	8	9	10	11	12
26	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.81
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.34
120	.10	.275	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	2.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	*6.76*	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
oo	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

TABLA C.5 Valores críticos de r (coeficiente de correlación producto-momento de Pearson)

	Nivel de significación para una prueba de dos colas*		
gl	.10	.05	.01
1	.988	.997	.9999
2	.900	.950	.990
3	.805	.878	.959
4	.729	.811	.917
5	.669	.754	.874
6	.622	.707	.834
7	.582	.*666*	.798
8	.549	.632	.765
9	.521	.602	.735
10	.497	.576	.708
11	.476	.553	.684
12	.458	.532	.661
13	.441	.514	.641
14	.426	.497	.623
15	.412	.482	.606
16	.400	.468	.590
17	.389	.456	.575
18	.378	.444	.561
19	.369	.433	.549
20	.360	.423	.537
25	.323	.381	.487
30	.296	.349	.449
35	.275	.325	.418
40	.257	.304	.393
45	.243	.288	.372
50	.231	.273	.354
60	.211	.250	.325
70	.195	.232	.303
80	.183	.217	.283
90	.173	.205	.267
100	.164	.195	.254

*Para una prueba de una cola, el nivel de significación se divide entre dos.

Apéndice D
Construcción de un cuadrado latino

Un cuadrado latino para determinar el orden de cualquier número *N* de condiciones, tiene *N* ordenamientos. Por lo tanto, si hay cuatro condiciones, hay cuatro ordenamientos en un cuadrado latino de 4 x 4; ocho condiciones producen un cuadrado latino de 8 x 8. El método para crear un cuadrado latino, que se muestra a continuación, produce ordenamientos en los que: 1) cada condición o grupo aparece sólo una vez en cada orden y 2) cada condición precede y sigue a cada condición en una ocasión.

Utilice el siguiente procedimiento para elaborar un cuadrado latino cuando haya una cantidad par de condiciones:

1. Determine el número de condiciones. Utilice letras del alfabeto para representar las *N* condiciones: ABCD para cuatro condiciones, ABCDEF para seis condiciones, etcétera.
2. Determine el orden del primer renglón por medio del siguiente ordenamiento:

   A, B, L, C, *L* - 1, D, *L* - 2, E

   y así sucesivamente. *L* se refiere al último tratamiento. Por lo tanto, si tiene cuatro condiciones (ABCD), el orden será

   A, B, D, C

   Con seis condiciones (ABCDEF), el orden será

   A, B, F, C, E, D

   debido a que Fes el último tratamiento (L) y E es el penúltimo (L - 1).

3. Determine el orden del segundo renglón incrementando una letra a cada posición del primer renglón. Por supuesto, la última letra no puede aumentarse, por lo que se reinicia con la primera letra. Con seis condiciones, el orden del segundo renglón es

   B, C, A, D, F, E

4. Continúe con este procedimiento para el tercer renglón y los subsecuentes. Para el tercer renglón, incremente una letra a cada posición del segundo renglón:

   C, D, B, E, A, F

   El cuadrado latino de 6 x 6 final es

A	B	F	C	E	D
B	C	A	D	F	E
C	D	B	E	A	F
D	E	C	F	B	A
E	F	D	A	C	B
F	A	E	B	D	C

5. Asigne aleatoriamente una letra a cada una de las condiciones para determinar qué condición estará en la posición A, en la posición B, y así sucesivamente.

   Si se trata de un número impar de condiciones, debe crear dos cuadrados latinos. Para el primer cuadrado, sólo siga el procedimiento anterior. Después elabore un segundo cuadrado que sea el inverso del primero; es decir, en cada renglón, la primera condición se convierte en la última, la siguiente condición en la penúltima, y así sucesivamente. Una los cuadrados para crear el cuadrado latino final (¡que en realidad es un rectángulo!). Así, si hay cinco condiciones, tendrá diez ordenamientos posibles para realizar su estudio.

Glosario

aleatorización Control de los efectos de variables extrañas al asegurarse que las variables operan de una forma totalmente determinada por el azar.

alfa de Cronbach Indicador de la confiabilidad de consistencia interna, que se evalúa cuando se examina la correlación promedio de cada reactivo (pregunta) en una medida, con cada una de las otras preguntas del mismo grupo de reactivos.

análisis de caso negativo En la observación de campo, un examen de las observaciones que no se ajustan a la estructura explicativa diseñada por el investigador.

análisis de contenido Análisis sistemático del contenido de registros escritos.

análisis de covarianza Técnica estadística para controlar la correlación entre una variable de sujeto y una variable dependiente en un experimento. Este procedimiento elimina la varianza de error que resulta del hecho de que la variabilidad en los puntajes de la variable dependiente se deben, en parte, al efecto de la variable de sujeto.

análisis de ruta Método utilizado para desarrollar modelos de posibles relaciones entre un conjunto de variables que fueron estudiadas con el método no experimental.

análisis de varianza Véase prueba F.

autonomía (Reporte Belmont) Principio que establece que los individuos pueden tomar la decisión de participar o no en una investigación.

beneficio (Reporte Belmont) Principio que establece que la investigación debe representar algunas ventajas para los participantes y minimiza cualquier consecuencia dañina.

características de demanda Claves que informan al sujeto sobre la forma en que se espera que se comporte.

coeficiente de correlación índice de la fuerza de la relación entre dos variables.

coeficiente de correlación producto-momento de Pearson Tipo de coeficiente de correlación utilizado con datos de escala de intervalo y de razón. Además de proporcionar información sobre la fuerza de la relación entre dos variables, el coeficiente de correlación producto-momento de Pearson indica la dirección (positiva o negativa) de la relación.

cohorte Grupo de personas que pertenecen a la misma generación (por nacimiento, ingreso a una institución, etc.) expuestas a los mismos sucesos sociales; los efectos de la cohorte pueden confundirse con la edad en un estudio transversal.

confederado (cómplice) La persona que finge ser participante en un experimento, pero que en realidad es aliado del investigador.

confiabilidad Grado de consistencia de una medida.

confiabilidad de consistencia interna Confiabilidad que se evalúa con los datos reunidos en un punto del tiempo, con múltiples medidas de un constructo psicológico. Una medida es confiable cuando las múltiples medidas ofrecen resultados similares.

confiabilidad entre jueces índice de confiabilidad que examina el grado de acuerdo entre las

observaciones realizadas por dos o más evaluadores (jueces).

confiabilidad por mitades Coeficiente de confiabilidad determinado por la correlación entre los puntajes de la mitad de los reactivos de un instrumento, con los puntajes de la otra mitad.

confiabilidad tes-tre-test Coeficiente de confiabilidad determinado por la correlación entre los puntajes de una medida en cierto momento, con los puntajes de la misma medida en un momento posterior.

conjunto de respuestas Patrón de respuesta de un individuo a las preguntas en una medida de autorreporte, que no está relacionado con el contenido de las preguntas.

consentimiento informado En la ética de la investigación, principio que establece que se debe informar de antemano a los participantes en un experimento sobre todos los aspectos del estudio que podrían influir en su decisión para participar.

contrabalanceo Método para controlar los efectos del orden en un diseño de medidas repetidas, al incluir todos los órdenes de presentación de los tratamientos o al determinar aleatoriamente el orden de los mismos para cada sujeto.

correlación múltiple Correlación entre una variable y una combinación de variables predictoras.

correlación parcial Correlación entre dos variables, con la influencia de una tercera variable estadísticamente controlada.

covariación de causa y efecto Parte de la inferencia causal; observar que un cambio en una variable independiente está acompañado por un cambio en una variable dependiente.

cuadrado latino Diseño empleado para controlar los efectos de orden de un diseño experimental, sin incluir todos los órdenes posibles.

definición operacional Definición de una variable que especifica las operaciones utilizadas para medirla o manipularla.

desviación estándar Desviación promedio de las puntuaciones respecto a la media (la raíz cuadrada de la varianza).

diferencias de selección Diferencias en el tipo de sujetos que conforman cada grupo en un diseño experimental; esta situación ocurre, por ejemplo, cuando los participantes eligen el grupo al que serán asignados.

diseño cuasiexperimental Tipo de diseño que se aproxima a las características de control de los experimentos verdaderos para deducir que una intervención dada tuvo el efecto esperado.

diseño de grupo con pretest y postest Diseño cuasiexperimental en el que el efecto de una variable independiente se infiere de la diferencia entre el pretest y el postest de un solo grupo.

diseño de grupo control no equivalente Diseño cuasiexperimental en el que grupos no equivalentes de sujetos participan en diferentes condiciones experimentales y donde no existe un pretest.

diseño de grupo únicamente con postest Diseño cuasiexperimental que carece de grupo control y de pretest de comparación; diseño muy pobre en términos de validez interna.

diseño de grupos apareados Forma de asignación de sujetos a los grupos, en la que primero se igualan pares de participantes respecto a alguna característica y después se asignan aleatoria e individualmente a los grupos.

diseño de grupos independientes Diseño en el que se asignan diferentes sujetos a cada grupo. También se le llama diseño entre sujetos.

diseño de medidas repetidas Diseño en ei cual todos los sujetos se exponen a todos los tratamientos. También llamado diseño dentro de sujetos (in-tra sujetos) o (mtra grupos).

diseño de reversión (reversible) Diseño, generalmente de caso único, en el que el tratamiento se introduce después de un periodo de línea base y que luego se retira dando paso a un segundo periodo

de línea base. Puede extenderse con una segunda introducción del tratamiento.

diseño de series de control Extensión del diseño cuasiexperimental de series de tiempo interrumpidas, en el cual existe un grupo de comparación o un grupo control.

diseño de series de tiempo interrumpidas Diseño en el que la eficacia de un tratamiento se determina al examinar una serie de mediciones realizadas durante un largo periodo, antes y después de introducir el tratamiento. El tratamiento no se introduce en forma aleatoria.

diseño de sólo postest Diseño experimental en el que la variable dependiente se mide una sola vez, después de la manipulación de la variable independiente (postest).

diseño factorial Diseño en el que se emplean al menos dos variables independientes; todos los niveles de cada variable independiente se combinan con todos los niveles de las otras variables independientes. Un diseño factorial permite investigar los efectos principales e interacciones de forma separada de dos o más variables independientes.

diseño factorial mixto Diseño que incluye variables de grupos independientes (entre sujetos) y de medidas repetidas (dentro de sujetos).

diseño pretest-postest Diseño experimental en el que la variable dependiente se mide antes (pretest) y después (postest) de la manipulación de la variable independiente.

diseño pretest-postest de grupo control no equivalente Diseño cuasiexperimental en el que se utilizan grupos no equivalentes, pero un pretest permite la evaluación de la equivalencia y de los cambios pretest-postest.

diseño VI • VP Diseño factorial que incluye una variable independiente experimental (VI) y una variable no experimental del participante (VP).

diseños de línea base múltiple Observación del comportamiento antes y después de una manipulación bajo múltiples circunstancias (entre individuos diferentes, conductas diferentes o condiciones diferentes).

distribución de frecuencias Ordenamiento de un conjunto de puntajes, del más bajo al más alto, que indica el número de *veces* que se obtuvo cada uno.

ecuación de regresión Ecuación matemática que permite predecir una conducta cuando se *conoce* el puntaje de otra variable.

efecto de acarreo Problema que puede ocurrir en los diseños de medidas repetidas, si los efectos de un tratamiento continúan presentes cuando se aplica el siguiente.

efecto de historia Como amenaza a la validez interna de un experimento, se refiere a cualquier evento colateral que no forma parte de la manipulación y que podría ser responsable de los resultados.

efecto de maduración Como amenaza a la validez interna, la posibilidad de que cualquier cambio que ocurra de manera natural dentro del individuo sea responsable de los resultados.

efecto de orden En el diseño de medidas repetidas, el efecto que tiene el orden de la introducción de los tratamientos sobre la variable dependiente.

efecto de piso Falla de una medición para detectar una diferencia, debido a la gran dificultad de la tarea relacionada (véase también efecto de techo).

efecto de prueba Amenaza a la validez interna, en la que el responder a un pretest cambia la conducta sin que haya existido ningún efecto de la variable independiente.

efecto de techo Falla de una medición para detectar una diferencia, debido a la facilidad de la tarea relacionada (véase también efecto de piso).

efecto principal Efecto directo de una variable independiente sobre una variable dependiente.

efecto principal simple En un diseño factorial, el efecto de una variable independiente en relación con un nivel particular de otra variable independiente.

efectos de instrumentación Como amenaza a la validez interna, la posibilidad de que las características de la instrumentación sean responsables de los resultados.

electroencefalograma (EEG) Medida de la actividad eléctrica del cerebro.

electromiograma (EMG) Medida de la actividad eléctrica de los músculos, incluyendo la tensión muscular.

empirismo Uso de observaciones objetivas para responder una pregunta acerca de la naturaleza del comportamiento.

encuadre de muestreo Individuos o grupo de individuos en una población que podrían ser seleccionados para incluirse en la muestra.

error de medición Grado en que una medición se desvía del valor verdadero.

error Tipo I Decisión incorrecta al rechazar la hipótesis nula cuando es verdadera.

error Tipo II Decisión incorrecta al aceptar la hipótesis nula cuando es falsa.

escala de intervalo Escala de medición en la que los intervalos entre los valores de la escala son del mismo tamaño.

escala de razón Escala de medición en la que existe el cero absoluto, el cual indica la ausencia de la variable que se está midiendo. Una de sus implicaciones es que se pueden establecer proporciones numéricas (generalmente consiste de medidas físicas como el peso o medidas de tiempo como la duración o el tiempo de reacción).

escala nominal Escala de medición con dos o más categorías que no poseen propiedades numéricas (menor que, mayor que).

escala ordinal Escala en la que las categorías de medición se ordenan numéricamente a lo largo de un continuo.

estadística descriptiva Medidas estadísticas que describen el resultado de un estudio; incluye medidas de tendencia central (por ejemplo, la media), de dispersión (por ejemplo, la desviación estándar) y de correlación (por ejemplo, la r de Pearson).

estadística inferencial Estadística diseñada para determinar si los resultados de datos muéstrales pueden ser generalizados a una población.

estudio de caso Informe descriptivo del comportamiento, historia y otros factores relevantes acerca de un individuo específico.

estudio de panel En la investigación de encuestas, entrevistar a las mismas personas en dos o más momentos.

estudio piloto Estudio a pequeña escala que se realiza antes de llevar a cabo un experimento real; diseñado para probar y refinar procedimientos.

evaluación de programa Investigación diseñada para evaluar programas creados para producir ciertos cambios en una población específica (por ejemplo, reformas e innovaciones sociales).

experimento de campo Experimento realizado en un ambiente natural y no en un ambiente de laboratorio.

experimento de un solo sujeto Experimento en el que el efecto de la variable independiente se evalúa analizando los datos de un solo participante, o de cada participante en lo individual.

explicación alternativa Parte de una inferencia causal; causa potencial alternativa de una relación observada entre variables.

grados de libertad (gl) Concepto que se utiliza en pruebas de significación estadística; número de observaciones que varían libremente para producir un resultado conocido.

gráfica de barras Uso de barras para mostrar frecuencias de respuestas, porcentajes o medias, en dos o más grupos.

gráfica de pastel Representación gráfica de datos en la que se representan frecuencias o porcentajes como "rebanadas" de un pastel.

grupo placebo En investigación, generalmente farmacológica, grupo que recibe una sustancia inerte para evaluar el efecto psicológico de recibir un tratamiento.

hipótesis Declaración que asevera que dos o más variables están relacionadas entre sí.

hipótesis de investigación Hipótesis que establece que las variables investigadas están relacionadas en la población; esto es que, con base en los datos muéstrales, el efecto observado es verdadero en la población.

hipótesis nula Hipótesis con propósitos estadísticos, que establece que las variables que se investigan no están relacionadas en la población; que cualquier efecto observado, basado en los resultados muéstrales, se debe a un error aleatorio.

informe Explicación que se da a los participantes después de su participación en la investigación, sobre los propósitos de ésta.

interacción Efecto diferencial de una variable independiente sobre la variable dependiente, en relación con el nivel particular de otra variable independiente.

intervalo de confianza Intervalo de valores, dentro del cual hay un nivel dado de confianza (por ejemplo, 95 por ciento) en el que se ubica el valor poblacional.

investigación documental Uso de fuentes de información ya existentes para una investigación. Las fuentes incluyen registros estadísticos, archivos de encuestas y registros escritos, entre otros.

IRM funcional La imagen de resonancia magnética utiliza un imán para obtener imágenes de estructuras del cerebro. La imagen de resonancia magnética funcional (ftMR) proporciona información sobre la cantidad de actividad en diferentes estructuras cerebrales.

juego de roles Procedimiento para el estudio del comportamiento, en el que se pide a los individuos que indiquen cómo responderían ante una situación dada, en lugar de ser observados directamente en dicha situación.

justicia (Reporte Belmont) Principio que establece que todos los individuos y grupos deben tener un acceso justo y equitativo a los beneficios por su participación en la investigación, así como a los riesgos potenciales de la misma.

lineábase Registro de la conducta de un sujeto o sujetos durante un periodo de control, sin introducir la manipulación experimental. Generalmente se utiliza en diseños con uno o pocos sujetos.

magnitud del efecto Grado de asociación entre dos variables. En la investigación experimental, la magnitud del impacto de la variable independiente sobre la variable dependiente.

media Medida de tendencia central que se obtiene sumando los puntajes individuales y dividiendo el total entre el número de éstos.

mediana Medida de tendencia central; el puntaje intermedio que divide una distribución en dos mitades.

meta-análisis Conjunto de procedimientos estadísticos para combinar los resultados de varios estudios, con el fin de proporcionar una evaluación general de la relación entre variables.

método correlacional Véase método no experimental.

método experimental Método que se utiliza para determinar si existe una relación entre variables, en el que el investigador manipula la variable independiente y controla todas las demás variables por medio de la aleatoriedad o por control experimental directo.

método longitudinal Método de investigación en el que se observa a las mismas personas de forma repetida durante un tiempo relativamente largo; conceptualmente similar al diseño de medidas repetidas.

método no experimental Uso de la medición de variables para determinar si están relacionadas entre sí. También llamado método correlacional.

método secuencial Combinación de los diseños transversal v longitudinal, generalmente para estudiar el desarrollo.

método transversal Método de investigación del desarrollo en el que se estudian personas de distintas edades en un mismo momento; conceptual-mente similar a un diseño de grupos independientes.

moda (modo) Medida de tendencia central; el puntaje más frecuente en una distribución de datos.

modelo estructural Modelo de un patrón esperado de relaciones entre un conjunto de variables. El patrón propuesto se basa en una teoría sobre cómo las variables se relacionan entre sí de manera causal.

mortandad Pérdida de los sujetos que abandonan un experimento. La mortandad es una amenaza a la validez interna cuando su tasa está relacionada con la naturaleza de la manipulación experimental.

muestreo Proceso de selección de los miembros de una población que serán incluidos en una muestra.

muestreo accidental (de conveniencia) Selección de sujetos de forma accidental, generalmente con base en la disponibilidad y no para obtener una muestra representativa de la población; tipo de muestreo no probabilístico.

muestreo aleatorio estratificado Procedimiento de muestreo en el que la población se divide en estratos, seguido por un muestreo aleatorio de cada estrato.

muestreo aleatorio simple Procedimiento de muestreo en el que cada miembro de la población tiene las mismas posibilidades de ser incluido en la muestra.

muestreo no probabilístico Todos los procedimientos de muestreo en los que no se puede especificar *la probabilidad de que cualquier miembro* de la población sea incluido en la muestra.

muestreo por cuotas Procedimiento de muestreo en el que la muestra se elige de modo tal que refleje la composición numérica de varios subgrupos de la población. Para obtener la muestra se emplea una técnica de muestreo accidental.

muestreo por racimos Método de muestreo en el cual se identifican grupos y subgrupos (racimos) de individuos. Se *realiza,* un muestreo de los racimos y después se incluye en la muestra a todos los individuos de cada racimo seleccionado.

muestreo probabilístico Tipo de procedimiento de muestreo en el que se especifica la probabilidad de que cualquier miembro de la población sea incluido en la muestra.

observación naturalista Método descriptivo en el que las observaciones se realizan en un ambiente social natural. También llamada observación de campo.

observación sistemática Observaciones de una o más conductas específicas, hechas generalmente en un ambiente definido con precisión.

población Grupo definido de individuos del que se obtiene una muestra.

polígono de frecuencias Presentación gráfica de una distribución de frecuencias, en la que la frecuencia de cada puntaje se gráfica en el eje vertical y los puntos se conectan con líneas rectas.

potencia Probabilidad de rechazar correctamente la hipótesis nula.

precedencia temporal Parte de la inferencia causal; la causa precede al efecto en una secuencia temporal.

predicción Afirmación que establece lo que ocurrirá en una investigación en particular.

probabilidad Posibilidad de que ocurra un suceso dado (entre un conjunto específico de sucesos).

prueba *F* (análisis de varianza) Prueba estadística para determinar si dos o más medias son *significativamente diferentes. F es la razón de la* varianza sistemática entre la varianza del error.

prueba *t* Prueba de significación estadística utilizada para comparar diferencias entre medias.

psicobiografía Tipo de estudio de caso en el que se analiza la vida de un individuo con base en una teoría psicológica.

puntaje verdadero Puntaje real de un individuo en una variable medida, en oposición al puntaje que el sujeto obtuvo en la propia medición.

reactividad Problema de medición en el que el procedimiento de medición cambia la conducta que se está observando.

reactivos de relleno Reactivos incluidos en un cuestionario para ayudar a disfrazar el verdadero propósito del instrumento.

realismo experimental Grado en que la manipulación de la variable independiente ejerce un impacto e involucra a los participantes en un experimento.

realismo mundano Grado en el que la manipulación de la variable independiente se asemeja a los eventos que ocurren en el mundo real.

regresión a la media También llamada regresión estadística, principio que establece que los puntajes extremos de una variable en una primera medición, tienden a acercarse a la media cuando se realiza una segunda medición.

relación curvilínea Relación en la que los incrementos en los valores de la primera variable están acompañados por incrementos y decrementos en los valores de la segunda variable.

relación lineal negativa Relación en la que los incrementos en los valores de la primera variable se acompañan por decrementos en los valores de la segunda variable.

relación lineal positiva Relación en la que los incremento de los valores de la primera variable están acompañados por incrementos en los valores de la segunda variable.

replicación Repetición de un estudio para determinar si los resultados pueden duplicarse.

replicación conceptual Replicación de una investigación que utiliza distintos procedimientos para manipular o medir las variables, pero que involucra los mismos elementos conceptuales.

replicación directa Replicación de una investigación con el uso de los mismos procedimientos para manipular y medir las variables que fueron empleadas en la investigación original.

respuesta galvánica de la piel (RGP) Conductancia eléctrica de la piel, que cambia cuando ocurre la sudoración.

sensibilidad Capacidad de una medida para detectar diferencias entre grupos o condiciones.

sesgo del entrevistador Influencia intencional o no intencional, ejercida por un entrevistador, de modo tal que el comportamiento real o interpretado de los participantes es consistente con las expectativas del entrevistador.

sesgo del experimentador (efectos de expectativa) Cualquier influencia intencional o no intencional que el experimentador ejerce sobre los sujetos para confirmar la hipótesis de investigación.

significación estadística Rechazo de la hipótesis nula, cuando un resultado tiene una baja probabilidad de ocurrencia por mero azar (generalmente de 0.05 o menos).

sistema de codificación Conjunto de reglas que se utilizan para establecer categorías de observaciones.

tendencia central Número o valor único que describe el puntaje típico o central de un conjunto de medidas.

validez Véase validez de constructo, validez externa, validez interna.

validez aparente Grado en que el instrumento de medición parece medir con precisión una variable.

validez concurrente La validez de constructo de una medición se evalúa cuando se examina si los grupos difieren en las formas esperadas.

validez convergente Forma de medir la validez de constructo en la que se evalúa la magnitud de la relación de los puntajes de una medición con los de otras mediciones del mismo constructo o cons-tructos similares.

validez de constructo Grado en el que un instrumento de medición mide con exactitud el constructo teórico para el que fue diseñado.

validez de criterio Técnica para determinar la validez de constructo, que se basa en la evaluación de la relación entre los puntajes de la medida y un criterio o resultado.

validez discriminante Forma de medir la validez de constructo en la que se evalúa la magnitud en que los puntajes de la medida no están relacionados con los de medidas conceptualmente no relacionadas.

validez externa Grado en que los resultados de un experimento pueden ser generalizados.

validez interna Certeza de que los resultados de un experimento pueden ser atribuidos a la manipulación de la variable independiente y no a otra variable.

validez predictiva La validez de constructo de una medida se evalúa examinando la capacidad de la medida para predecir una conducta futura.

variabilidad Dispersión de los puntajes respecto al valor de alguna medida de tendencia central.

variable Cualquier suceso, situación, conducta o característica individual que varía; es decir, que tiene al menos dos valores.

variable de confusión Variable no controlada que varía sistemáticamente con una variable independiente. En un experimento, los grupos experimentales difieren tanto en la variable independiente como en la variable de confusión.

variable dependiente Variable que corresponde a alguna característica de la respuesta del sujeto, cuyos valores cambian con relación a los cambios en una variable independiente.

variable independiente Variable que se manipula para observar sus efectos sobre la variable dependiente.

variable moderadora Variable que influye en la naturaleza de la relación entre otras dos variables (una variable independiente y una variable dependiente). En un diseño factorial, el efecto de la variable moderadora se manifiesta como una interacción.

varianza Medida de la variabilidad de los puntajes respecto a la media; la media de la suma de las des-viaciones cuadradas de las puntuaciones respecto de la media grupal.

varianza de error Variabilidad aleatoria en un conjunto de puntuaciones, que no es el resultado de la variable independiente. Estadísticamente, la variabilidad de cada puntaje con respecto a su media grupal.

varianza sistemática Variabilidad en un conjunto de puntajes, que es resultado de la variable independiente; estadísticamente, la variabilidad de la media de cada grupo a partir de la gran media de todos los sujetos.

verificación de la manipulación Medida utilizada para determinar si la manipulación de la variable independiente produjo el efecto esperado en un participante.

Referencias

Aiello, J. R., Baum, A. y Gormley, F. P. (1981). Social determinants of residential crowding stress. *Personality and Social Psycbology Bulletin, 4,* 643-649.

Ajzen, I. y Fishbein, M. (1980). *Understanding altitudes and predicting social behavior* Englewood Cliffs, NJ: Prentice-Hall.

American Psychological Association. (2002a). *Ethi-cal principies of psychologists and code of conduct.* Consultado en octubre 20, 2002, de http://www.apa.org/ethics/code2002.html

American Psychological Association. (2002b). *Guidelines for ethical conduct in the care and use of animáis.* Consultado en marzo 2, 2003, de http://www.apa.org/science/anguide.html

American Psychological Association. (2001). *Publicaron manual of the American Psychological Association* (5a. ed.). Washington, DC: Autor.

Anderson, C. A. y Anderson, D. C. (1984). Ambient temperature and violent crime: Test of the linear and curvilinear hypotheses. *Journal ofPer-sonality and Social Psycbology, 46,* 91-97.

Anderson, C. A., Lindsay, J. J. y Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8}* 3-9.

Anderson, D. R., Lorch, E. R, Field, D. E., Collins, P. A. y Nathan, J. G. (1986). TV viewing at home. *Child Development, 57,* 1024-1033.

Aristotle. (1954). Rhetoric. En *Aristotle, rhetoric, and poetics* (W. Rhys Roberts, Trans.). Nueva York: Modern Library.

Aronson, E., Brewer, M. y Carlsmith, J. M. (1985). Experimentation in social psychology. En G. Lindzey y E. Aronson (Eds.), *Handbook of social psychology* (3a. ed.). Nueva York: Random House.

Asch, S. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 709* (Whole Núm. 416).

Astin, A. (1987). *The American freshman: Twentyyear trends, 1966-1985.* Los Angeles: Higher Education Research Institute, Gradúate School of Education, University of California.

Bakeman, R. y Brownlee, J. R. (1980). The strategic use of parallel play: A sequential analysis. *Child Development, 51,* 873-878.

Bakeman, R. y Gottman, J. M. (1986). *Observinginteraction.* Cambridge: Cambridge University Press.

Bales, R F. y Cohén, S. P. (1979). *SYMLOG: A system for the múltiple level observation of groups.* Nueva York: Free Press.

Barlow, D. H. y Hersen, M. (1984). *Single case experimental designs.* Nueva York: Pergamon Press.

Barón, R. A. (1997). The sweet smell of.. . helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. *Personality and Social Psychology Bulletin, 23,* 498-503.

Barón, R. M. y Kenney, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Barton, E. M., Baltes, M. M. y Orzech, M. J. (1980). Etiology of dependence in older nursing home residents during morning care; the role of staff behavior. *Journal of Personality and Social Psychology, 38,* 423-431.

Baum, A., Gachtel, R. J. y Schaeffer, M. A. (1983). Emotional, behavioral, and psychological effects of chronic stress at Three Mile lsland. *Jour-*

nal of Consulting and Clinical Psycbology, 51, 565-572.

Baumeisten R. F. v Xewman, L. S. (1994). How stories make sense of personal experiences: Motives that shape autobiographical narratives. *Personality and Social Psychology Bulletin, 20,* 676-690.

Beach, F. A. (1950). The snark was a boojum. *America n Psychologist, 5,* 115-124.

Becker, H. S. (1963). *Outsiders: Studies in the sociology of deviance.* Nueva York: Free Press.

Bem, D. J. (1981). Writing the research report. En L. H. Kidder (Ed.), *Research methods in social relations.* Nueva York: Holt, Rmehart y Winston.

Bem, D. J. (2002). *Writing the empirical journal árdele.* Consultado en agosto 15, 2002, de http:// comp9.psych.cornell.edu/dbem/writing_article.html

Berry, T D. y Geller, E. S. (1991). A single-subject approach to evaluative vehicle safety belt reminders: Back to basics. *Journal of Applied Bebavior Analysis, 24,* 13-22.

Berscheid, E., Barón, R. S., Dermer, M. y Libman, M. (1973). Anticipating informed consent: An empirical approach. *American Psychologist, 28,* 913-925.

Bornstein, B. H. (1998). From compassion to compensation: The effect of injury severity on mock jurors' liability judgments. *Journal of Applied Social Psycbology, 28,* 1477-1502.

Bortnik, K., Henderson, L. y Zimbardo, P. (2002). *Tbe Shy Q, a measure of chronic shyness: Associations with interpersonal motives and interpersonal valúes.* Consultado en noviembre 10, 2002, de http: //www.shyness.com/documents/2002/SITAR_ 2002poster_handout.pdf

Bouchard, T J., Jr. y McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055-1059.

Broberg, A. G., Wessels, H., Lamb, M. E. y Hwang, C. P. (1997). Effects of day care on the develop-ment of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology, 33,* 62-69.

Bróder, A. (1998). Deception can be acceptable. *American Psychologist, 53,* 805-806.

Brogden, W. J. (1962). The experimenter as a factor in animal conditioning. *Psycbological Reports, 11,* 239-242.

Brooks, C. I. y Rebata, J. L. (1991). College class-room ecology: The relation of sex of student to classroom performance and seating preference. *Environment and Bebavior, 23,* 305-313.

Brown, A. S. y Rahhal, T. A. (1994). Hidmg valuables: A questionnaire study of mnemonically risky behavior. *Applied Cognitive Psychology, 8,* 141-154.

Brunmg, N. S. y Frew, D. R. (1987). Effects of exercise, relaxation, and management skills training on physiological stress indicators: A field ex-periment. *Journal of Applied Psycbology, 72,* 515-521.

Bushman, B. J. y Wells, G. L. (2001). Narrative ímpressions of the literature: The availability bias and the corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin,27,* 1123-1130.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Bebavioral and Brain Sciences, 12,* 1-49.

Buss, D. M. (1998). *Evolutionary psychology: Tbe New science oftbe mind.* Boston: Allyn y Bacon.

Byrne, D., Ervm, C. R. y LamberthJ, (1970). Continuity between the experimental study of attraction and real-life computer dating. *Journal of Personality and Social Psycbology, 16,* 157-165.

Byrne, G. (1988, octubre 7). Breuning pleads guilty. *Science, 242,* 27-28.

Cacioppo, J. T. y Tassinary, L. G. (1990). Inferring psychological significance from physiological signáis. *American Psychologist, 45,* 16-28.

Campbell, D. T. (1968). Quasi-experimental de-sign. En D. L. Gillis (Ed.), *International encyclopedia of tbe social sciences* (Vol. 5). Nueva York: Macmillan and Free Press.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24,* 409-429.

Campbell, D. T. y Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Carroll, M. E. y Overmier, J. B. (Eds.). (2001). *Animal research and human health: Advancing human welfare through behavioral science.* Washington, DC: American Psychological Association.

Caspi, A. y Silva, P. T. (1995). Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development, 66,* 486-498.

Cavan, S. (1966). *Liquor license: An ethnography of bar behavior*. Chicago: Aldine.

Chaiken, S. y Pliner, P. (1987). Women, but not men, are what they eat: The effect of meal size and gender on perceived femininity and masculinity. *Personality and Social Psychology Bulletin, 13,* 166-176.

Chastain, G. D. y Landrum, R. E. (Eds.). (1999). *Protecting human subjects: Department subject pools and institutional review hoards.* Washington, DC: American Psychological Association.

Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin, 14, 664-675.*

Cialdini, R. B. (1988). *Influence: Science and practice* (2a. ed.). Glenview, IL: Scott, Foresman.

Clark, K. B. y Clark, M. P. (1947). Racial identification and preference in Negro children. En T. M. Newcomb y E. L. Hartley (Eds.), *Readings in social psychology*. Nueva York: Holt, Rinehart y Winston.

Cohén, D., Nisbett, R. E., Bowdle, B. F. y Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An "experimental ethnography." *Journal of Personality and Social Psychology, 70,* 945-960.

Cohén, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohén, J. (1994). The earth is round *(p,* .05). *American Psychologist, 49,* 997-1003.

Coile, D. C. y Miller, N. E. (1984). How radical animal activists try to mislead humane people. *American Psychologist, 39,* 700-701.

Coltheart, V. y Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory y Cognition, 26,* 330-342.

Conoley, J. C. y Kramer, J. J. (1989). *Tenth mental measurements yearbook.* Lincoln: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Converse, J. M. y Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire.* Newbury Park, CA: Sage.

Cooper, H. M. y Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87,* 442-449.

Costa, P. T, Jr. y McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Cunningham, M. R., Druen, P. B. y Barbee, A. P. (1997). Angels, mentors, and friends: Trade-offs among evolutionary, social, and individual variables in physical appearance. En J. A. Simpson y D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 109-140). Mahwah, NJ: Erlbaum.

Cunningham, M. R., Roberts, R., Barbee, A. R, Druen, P. B. y Wu, C. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68,* 261-279.

Curtiss, S. R. (1977). *Genie: A psycholinguistic study of a modern-day "wild child."* Nueva York: Academic Press.

Darley, J. M. y Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377-383.

Denmark, F., Russo, N. P, Frieze, I. H. y Sechzer, J. A. (1988). Guidelines for avoiding sexism in psychological research: A report of the Ad Hoc Committee on Nonsexist Research. *American Psychologist, 43,* 582-585.

Department of Health and Human Services. (2001). Protection of human subjects. Consultado en marzo 19,2003, de http:/ohrp. osophs. dhhs.gov/humansubjects/guidance/45cfr46. htm

Dermer. M. L. y Hoch, T. A. (1999). Improving descripnons of single-subject experiments in re-search texts written for undergraduates. *Psycho-logical Record, 49,* 49-66.

Dill, C. A.; Gilden, E. R, Hill, P. C. y Hanslka, L. L. (1982). Federal human subjects regulations: A methodological artifact? *Personality and Social Psychology Bulletin, 8,* 417-425.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2a. ed.). Nueva York: Wiley.

Duncan, S., Rosenberg, M. J. y Finklestein, J. (1969). The paralanguage of experimenter bias. *Sociometry, 32,* 207-219.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G. y Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psycho-logical Bulletin, 110,* 109-128.

Eagly, A. M. y Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist, 54,* 408-423.

Ellsworth, P. C., Carlsmith, J. M. y Henson, A. (1972). The stare as a stimulus to flight in human beings: A series of field experiments. *Journal of Personality and Social Psychology, 21,* 302-311.

Elms, A. C. (1994). *Uncovering Uves: The uneasy alliance of biography and psychology.* Nueva York: Oxford University Press.

Ennett, S. T, Tobler, N. S., Ringwalt, C. L. y Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project D.A.R.E. outcome evaluations. *American Journal of Public Health, 84,* 1394-1401.

Epstein, Y. M., Suedfeld, P. y Silverstein, S. J. (1973). The experimental contract: Subjects' expectations of and reactions to some behaviors of experimenters. *American Psychologist, 28,* 212-221.

Everett, R B., Hayward, S. C. y Meyers, A. W. (1974). The effects of a token reinforcement procedure on bus ridership. *Journal of Applied Behavior Analysis, 7,* 1-10.

Fazio, R. H., Cooper, M., Dayson, K. y Johnson, M. (1981). Control and the coronary-prone behavior pattern: Responses to múltiple situational demands. *Personality and Social Psychology Bulletin, 7,* 97-102.

Feldman Barrett, L. y Barrett, D. J. (2001). Computerized experience-sampling: How technology facilitates the study of conscious experience. *Social Science Computer Review, 19,* 175-185.

Fischer, K., Schoeneman, T. J. y Rubanowitz, D. E. (1987). Attributions in the advice columns: II. The dimensionality of actors' and observers' explanations of interpersonal problems. *Personality and Social Psychology Bulletin, 13,* 458-466.

Fiske, S. T, Bersoff, D. N., Borgida, E., Deaux, K. y Heilman, M. E. (1991). Social science research on trial: Use of sex stereotyping in Price Waterhouse v. Hopkins. *American Psychologist, 46,* 1049-1060.

Fiske, S. T y Taylor, S. E. (1984). *Social cognition.* Nueva York: Random House.

Flavell, J. H. (1996). Piaget's legacy. *Psychological Science, 7,* 200-203.

Fowler, R J., Jr. (1984). *Survey research methods.* Newbury Park, CA: Sage.

Frank, M. G. y Gilovich, T (1988). The dark side of self- and social perception: Black uniforms and aggression in professional *sports. Journal of Personality and Social Psychology, 54,* 74-85.

Freedman,J. L. (1969). Role-playing: Psychology by consensus. *Journal of Personality and Social Psychology, 13,* 107-114.

Freedman, J. L., Klevansky, S. y Ehrlich, P. R. (1971). The effect of crowding on human task performance. *Journal of Applied Social Psychology, 1,* 7-25.

Freedman, J. L., Levy, A. S., Buchanan, R. W. y Price, J. (1972). Crowding and human aggressiveness. *Journal of Experimental Social Psychology, 8,* 528-548.

Frick, R. W. (1995). Accepting the null hypothesis. *Memory and Cognition, 25,* 132-138.

Fried, C. B. (1999). Who's afraid of rap: Differen-tial reactions to music lyrics. *Journal of Applied Social Psychology, 29,* 705-721.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Tomlinson-Keasy, C, Martin, L. R., Wingard, D.

L. y Criqui, M. H. (1995). Psychosocial and behavioral predictors of longevity: The aging and death of the "Termites." *American Psychologist, 50,* 69-78.

Furnham, A., Gunter, B. y Peterson, E. (1994). Televisión distraction and the performance of in-troverts and extroverts. *Applied Cognitive Psychology, 8,* 705-711.

Gallup, G. G. y Suarez, S. D. (1985). Alternatives to the use of animáis in psychological research. *American Psycbologist, 40,* 1104-1111.

Gardner, G. T. (1978). Effects of federal human subjects regulations on data obtained in environmental stressor *research. Journal of Personality and Social Psychology, 34,* 774-781.

Gardner, L. E. (1988). A relatively painless method of introduction to the psychological literature search. En M. E. Ware y C. L. Brewer (Eds.), *Handbook for teaching statidistics and research methods.* Hillsdale, NJ: Erlbaum.

Gaulm, S. J. C. y McBurney, D. (2000). *Psychology: An evolutionary approach.* Upper Saddle River, NJ: Prentice-Hall.

Gelfand, H. y Walker, C. J. (2001). *Mastering APA style.* Washington, DC: American Psychological Association.

Geller, E. S., Russ, N. W. y Ahornan, M. G. (1986). Naturalistic observations of beer drinkmg among college students. *Journal of Applied Behavior Analysis, 19,* 391-396.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* Nueva York: Free Press.

Goodstem, D. (2000). *How science works.* Consultado en febrero 10, 2002, de http://www.its. caltech.edu/dg/HowScien.pdf

Graesser, A. C., Kennedy, T, Wiemer-Hastings, P. y Ottati, V. (1999). The use of computational cognitive methods to improve questions on surveys and questionnaires. En M. G. Sirkin, D. J. Hermann, S. Schechter, N. Schwarz, J. M. Tanur y R. Tourangeau (Eds.), *Cognition and survey methods research* (pp. 199-216). Nueva York: Wiley.

Green, j. y Wallaf, C. (1981). *Ethnography and language in educational settings.* Nueva York: Ablex.

Greenfield, D. N. (1999). *Nature of Internet addic-tion: Psychological factors in compulsive Internet use.* Documento presentado en la American Psychological Association, Boston, MA.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin, 83,* 314-320.

Gross, A. E. y Fleming, I. (1982). Twenty years of deception in social psychology. *Personality and Social Psychology Bulletin, 8,* 402-408.

Gwaltney-Gibbs, P. A. (1986). The institutionalization of premarital cohabitation: Estimates from marriage license applications, 1970 and 1980. *Journal of Marriage and the Family, 48,* 423-434.

Haney, C. y Zimbardo, P. G. (1998). The past and future of U.S. prison policy: Twenty-five years after the Stanford Prison Experiment. *American Psycbologist, 53,* 709-727.

Harris, R. (2002). *Antiplagiarism strategies for research papers.* Consultado en septiembre 10,2002, de http://wuAv.virtualsalt.com/antiplag.htm

Hawking, S. W. (1988). *A hrief history of time: From the big bang to black boles.* Nueva York: Bantam Books.

Hearnshaw, L. S. (1979). *Cyril Burt, psychologist.* Ithaca, NY: Cornell University Press.

Henle, M. y Hubbell, M. B. (1938). "Egocentncity" in adult conversation. *Journal of Social Psychology, 9,* 227-234.

Hermán, D. B., Struening, E. L. y Barrow, S. M. (1994). Self-reported needs for help among homeless men and women. *Evaluation and Program Planning 17,* 249-256.

Hill, C. T, Rubín, Z. y Peplau, L. A. (1976). Break-ups before marriage: The end of 103 affairs. *Journal of Social Issues, 32,* 147-168.

Hill, L. (1990). Effort and reward m college: A replication of some puzzling findings. En J. W. Neuliep (Ed.), *Handbook of replication in the behavioral and social sciences* [Special issue]. *Journal of Social Behavior and Personality,* 5(4), 151-161.

Holden, C. (1987). Animal regulations: So far, so good. *Science, 238,* 880-882.

Hood, T. C. y Back, K. W. (1971). Seif-disclosure and the volunteer: A source of bias in labora-

tory experiments. *Journal of Personality and Social Psycbology, 17,* 130-136.

Hostetler, A. J. (mayo de 1987). Fraud inquiry revives doubt: Can science police itself? *APA Monitor,* 1, 12.

Humphreys, L. (1970). *Tearoom trade.* Chicago: Aldine.

Jones, R. y Cooper, J. (1971). Mediation of experimenter effects. *Journal of Personality and Social Psycbology, 20,* 70-74.

Josselson, R. y Lieblich, A. (1993). *The narrative study oflives.* Newbury Park, CA: Sage.

Jourard, S. M. (1969). The effects of experimenters' self-disclosure on subjects' behavior. En C. Spielberger (Ed.), *Current topics in community and clinical psycbology.* Nueva York: Academic Press.

Joy, L. A., Kimball, M. M. y Zabrack, M. L. (1986). Television and children's aggressive behavior. En T. M. Williams (Ed.), *The impact of television: A natural experiment in tbree communities.* Orlando, FL: Academic Press.

Judd, C. M., Smith, E. R. y Kidder, L. H. (1991). *Research methods in social relations* (6a. ed.). Ft. Worth, TX: Holt, Rinehart y Winston.

Kamin, L. G. (1974). *The science and politics of IQ.* Nueva York: Wiley.

Kazdin, A. E. (1995). Preparing and evaluacing research reports. *Psychological Assessment, 7,* 228-237.

Kazdin, A. E. (2001). Behavior modification m applied settings (6a. ed.). Belmont, CA: Wádsworth.

Kelman, H. C. (1967). Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin, 67,* 1-11.

Kenny, D. A. (1979). *Correlation and causality.* Nueva York: Wiley.

Keyser, D. J. y Sweetland, R. C. (Eds.). (1991). *Test critiques.* Kansas City, MO: Test Corporation of America.

Kimmel, A. (1998). En defense of deception. *Amencan Psycbologist, 53,* 803-805.

Kintz, N. L., Delprato, D. J., Mettee, D. R, Persons, C. E. y Schappe, R. H. (1965). The experimenter effect. *Psychological Bulletin, 63,* 223-232.

Kirsch, L, Moore, T. J., Scoboria, A. y Nicholls, S. S. (2002). The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention y Treatment, 5,* Artículo 23. Consultado en noviembre 1, 2002, de http:// www. journals. apa. org/prevention/volume5/ pre0050023a.html

Kitayama, S., Markus, H. R., Matsumoto, H. y Norasakkunkit, V. (1997). Endividual and *co*llective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology, 72,* 1245-1267.

Koocher, G. P. (1977). Bathroom behavior and human dignity. *Journal of Personality and Social Psychology, 35,* 120-121.

Koop, C. E. (1987). Report of the Surgeon General's workshop on pornography and public health. *American Psycbologist, 42,* 944-945.

Korn, J. H. (1997). *Illusions of reality: A history of deception in social psychology.* Albany: State University of Nueva York Press.

Korn, j. H. (1998). The reality of deception. *American Psycbologist, 53,* 805.

Koss, M. P. (1992). The underdetection of rape: Methodololgical choices influence incident estimates. *Journal of Social Issues, 48(1),* 61-75.

Krantz, J. H., Bailará, J. y Scher, J. (1997). Comparing the results of laboratory and World Wide Web samples on the determinants of female attractiveness, *Behavior Research Methods, Instrumentation, and Computer, 29,* 264-269.

Labranche, E. R., Helweg-Larsen, M., Byrd, C. E. y Choquette, R. A., Jr. (1997). To picture or not to picture: Levels of erotophobia and breast self-examination brochure techniques. *Journal of Applied Social Psycbology, 27,* 2200-2212.

Lana, R. E. (1969). Pretest sensitization. En R. Rosenthal y R. L. Rosnow (Eds.), *Artifacts in behavioral research.* Nueva York: Academic Press.

Langer, E.J. y Abelson, R. P. (1974). A patient by any other name . . . : Clinical group difference in labeling bias. *Journal of Consulting and Clinical Psychology, 42,* 4-9.

Langer, E. J. y Rodin, J. (1976). The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional set-ting. *Journal of Personality and Social Psychology, 34,* 191-198.

Latané, B., Williams, K. y Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37,* 822-832.

Leventhal, H. (1970). Findings and theory in the study of fear Communications. En L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5). Nueva York: Academic Press.

Levin, J. R. (1983). Pictorial strategies for school learning: Practical illustrations. En M. Pressley y J. R. Levin (Eds.), *Cognitive strategy research: Educational applications* (pp. 213-238). Nueva York: Springer-Verlag.

Levine, R. V. (1990). The pace of life. *American Scientist, 78,* 450-459.

Linz, D., Donnerstein, E. y Penrod, S. (1987). The findings and recommendations of the Attorney GeneraPs Commission on Pornography: Do the psychological "facts" fit the political fury? *American Psychologist, 42,* 946-953.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis.* Mahwah, NJ: Erlbaum.

Lofland, J. y Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis* (3a. ed.). Belmont, CA: Wadsworth.

Loftus, E. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Luria, A. R. (1968). *The mind of a mnemonist.* Nueva York: Basic Books.

Macintosh, N. J. (Ed.). (1995). *Cyril Burt: Fraud or framed?* Nueva York: Oxford University Press.

Marlatt, G. A. y Rohsenow, D. R. (1980). Cognitive processes in alcohol use: Expectancy and the balanced placebo design. En N. K. Mello (Ed.), *Advances in substance abuse* (Vol. 1). Greenwich, CT: JAI Press.

Maruyama, G. M. (1998). *Basics of structural equation modeling.* Thousand Oaks, CA: Sage.

Matsumoto, D. (1994). *Cultural influences on research methods and statistics.* Belmont, CA: Brooks/ Colé.

Matteson, M. T. e Ivancevich, J. M. (1983). Note on tensión discharge rate as an employee health status predictor. *Academy of Management Journal, 26,* 540-545.

Mazzella, R. y Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24,* 1315-1344.

McCutcheon, L. E. (2000). Another failure *to* generalize the Mozart effect. *Psychological Reports, 87,* 325-330.

McGuigan, F. J. (1963). The experimenter: A neglected stimulus. *Psychological Bulletin, 60,* 421-428.

Middlemist, R. D., Knowles, E. S. y Matter, C. F. (1976). Personal space invasion in the lavatory: Suggestive evidence for arousal. *Journal of Personality and Social Psychology, 33,* 541-546.

Middlemist, R. D., Knowles, E. S. y Matter, C. F. (1977). What to do and what to report: A reply *to* Koocher. *Journal of Personality and Social Psychology, 35,* 122-124.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371-378.

Milgram, S. (1964). Group pressure and action against a person. *Journal of Abnormal and Social Psychology, 69,* 137-143.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18, 57-76.*

Miller, A. G. (1972). Role-playing: An alternative to deception? *American Psychologist, 27,* 623-636.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science.* Nueva York: Praeger.

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist, 24,* 1063-1075.

Miller, J. G. (1999). Cultural psychology: Implications for basic psychological theory. *Psychologi-cal Science, 10,* 85-91.

Miller, N. E. (1985). The valué of behavioral re-search on animáis. *American Psychologist, 40,* 423-440.

Montee, B. B., Miltenberger, R. G. y Wittrock, D. (1995). An experimental analysis of facilitated communication. *Journal of Applied Behavior Analysis, 28,* 189-200.

Morgan, D. L. y Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist, 56,* 119-127.

Murray, B. (2002). Research fraud needn't happen at all. *APA Monitor, 33(2).* Consultado en julio 31, 2002, de http://www.apa.org/monitor/feb02/ fraud.html

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (April 18, 1979). *The Belmont Report: Ethical principies and guidelines for the protection of human subjects of research.* Consultado en marzo 19, 2003, de http://ohsr.od.nih.gov/mpa/ belmont. php3

Nicol, A. A. M. y Pexman, P. M. (2003). *Displaying your findings: A practical guide for creating figures, posters, andpresentations.* Washington, DC: American Psychological Association.

Nisbett, R. E. y Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E. y Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Ono, H., Phillips, K. A. y Leneman, M. (1996). Contení of an abstract: De jure and de facto. *American Psychologist, 51,* 1338-1340.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference *to* demand characteristics and their implications. *American Psychologist, 17,* 776-783.

Ortmann, A. y Hertwig, R. (1997). Is deception acceptable? *American Psychologist, 52,* 746-747.

Osgood, C. E., Suri, G. J. y Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Patterson, G. R. y Moore, D. (1979). Interactive patterns as units of behavior. En M. E. Lamb, S. J. Sumoi y G. L. Stephenson (Eds.), *Social interaction analysis: Methodological issues* (pp. 77-96). Madison: University of Wisconsin Press.

Paulus, P. B., Annis, A. B., Seta, J. J., Schkade, J. K. y Matthews, R. W. (1976). Crowding does affect task *performance. Journal of Personality and Social Psychology, 34,* 248-253.

Pepitone, A. y Triandis, H. (1987). On the universality of social psychological theories. *Journal of Cross-Cultural Psychology, 18,* 471-499.

Petty, R. E. y Cacioppo, J. T. (1986). *Communication and persuasión: Central and peripheral routes to attitude change.* Nueva York: Springer-Verlag.

Petty, R. E., Cacioppo, J. T. y Goldman, R. (1981). Personal involvement as a determinant of argumentbased persuasión. *Journal of Personality and Social Psychology, 41,* 847-855.

Pfungst, O. (1911). *Clever Hans (the horse ofMr. von Osten): A contribution to experimental, animal, and human psychology* (C. L. Rahn, Trans.). Nueva York: Holt, Rinehart y Winston. (Edición de 1965.)

Piaget, J. (1952). *The origins of intelligence in children.* Nueva York: International Universities Press.

Plous, S. (1996a). Attitudes toward the use of animáis in psychological research and education: Results from a national survey of psychologists. *American Psychologist, 51,* 1167-1180.

Plous, S. (1996b). Attitudes toward the use of animáis in psychological research and education: Results from a national survey of psychology majors. *Psychological Science, 7,* 352-363.

Punnett, B. J. (1986). Goal setting: An extensión of the research. *Journal of Applied Psychology, 71,* 171-172.

Rauscher, F. H. y Shaw, G. L. (1998). Key components of the Mozart effect. *Perceptual and Motor Skills, 86,* 835-841.

Rauscher, F. H, Shaw, G. L. y Ky, K. N. (1993). Music and spatial task performance. *Nature, 365,* 611.

Rauscher, F. H, Shaw, G. L. y Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters, 185,* 44-47.

Reed, J. G. y Baxter, P. M. (2003). *Library use: A bandbook for psychology* (3a. ed.). Washington, DC: American Psychological Association.

Reeve, D. K. y Aggleton, J. P. (1998). On the specificity of expert knowledge about a soap opera: An everyday story of farming folk. *Applied Cognitive Psycbology, 12,* 35-42.

Reifman, A. S, Larrick, R. P. y Fein, S. (1991). Temper and temperature on the diamond: The heat-aggression relationship in major league baseball. *Personality and Social Psychology Bulletin, 17,* 580-585.

Reverby, S. M. (Ed.). (2000). *Tuskegees trutbs: Retbinking the Tuskegee sypbilis study.* Chapel Hill, NC: University of North Carolina Press.

Ring, K. (1967). Experimental social psychology: Some sober questions about frivolous *values. Journal of Experimental Social Psychology, 3,* 113-123.

Ring, K., Wallston, K. y Corey, M. (1970). Mode of debriefing as a factor affecting subjective reaction to a Milgram-type obedience experiment: An ethical inquiry. *Representative Research in Social Psychology, 1,* 67-68.

Riordan, C. A. y Marlin, N. A. (1987). Some good news about some bad practices. *American Psychologist, 42,* 104-106.

Roberson, M. T. y Sundstrom, E. (1990). Questionnaire design, return rates, and response favorableness in an employee attitude questionnaire. *Journal of Applied Psychology, 75,* 354-357.

Robinson, J. P, Athanasiou, R. y Head, K. B. (1969). *Measures of occupational attitudes and occupational characteristics.* Ann Arbor, MI: Institute for Social Research.

Robinson, J. P, Rusk, J. G. y Head, K. B. (1968). *Measures of political attitudes.* Ann Arbor, MI: Institute for Social Research.

Robinson, J. P, Shaver, P R. y Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes* (Vol. 1). San Diego, CA: Academic Press.

Rodin, J. y Langer, E. J. (1977). Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of Personality and Social Psychology, 35,* 897-902.

Rosenbaum, D. P y Hanson, G. S. (1998). Assessing the effects of school-based drug education: A six-year multilevel analysis of Project D.A.R.E. *Journal of Research in Crime and Delinquency, 35,* 381-412.

Rosenblatt, P C. y Cozby, P C. (1972). Courtship patterns associated with freedom of choice of spouse. *Journal of Marriage and the Family, 34,* 689-695.

Rosenhan, D. (1973). On being sane in insane places. *Science, 179,* 250-258.

Rosenthal, R. (1966). *Experimenter effects in behavior research.* Nueva York: Appleton-Century-Crofts.

Rosenthal, R. (1967). Covert communication in the psychological experiment. *Psychological Bulletin, 67,* 356-367.

Rosenthal, R. (1969). Interpersonal expectations: Effects of the experimentéis hypothesis. En R. Rosenthal y R. L. Rosnow (Eds.), *Artifacts in behavioral research.* Nueva York: Academic Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rosenthal, R. yjacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupiW intellectual development.* Nueva York: Holt, Rinehart & Winston.

Rosenthal, R. y Rosnow, R. L. (1975). *The volunteer subject.* Nueva York: Wiley.

Rosnow, R. L. y Rosnow, M. (2002). *Writingpapers in psychology* (6a. ed.). Belmont, CA: Wadsworth.

Rossi, P H., Freeman, H. E. y Lipsey, M. W. (1999). *Evaluation: A systematic approach.* Thousand Oaks, CA: Sage.

Ruback, R. B. y Juieng, D. (1997). Territorial de-fense in parking lots: Retaliation against wai-ting drivers. *Journal of Applied Social Psychology, 27,* 821-834.

Rubin, Z. (diciembre de 1970). Jokers wild in the lab. *Psychology Today, 18,* 20, 22-24.

Rubin, Z. (1973). Designing honest experiments. *American Psychologist, 28,* 445-448.

Rubín, Z. (1975). Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology, 11,* 233-260.

Rubín, Z. (1985). Deceiving ourselves about deception: Comment on Smith and Richardson's "Amelioration of deception and harm in psychological *research." Journal of Personality and Social Psychology, 48,* 252-253.

Russell, C. H. y Megaard, I. (1988). *The general social survey, 1972-1986: The state of the American people.* Nueva York: Springer-Verlag.

Russell, D., Peplau, L. A. y Cutrona, C. E. (1980). The revised UCLA Loneliness Scale: Concu-rrent and discriminant validity. *Journal of Personality and Social Psychology, 39,* 472-480.

Saks, M. J. y Marti, M. W. (1997). A meta-analysis of the effects of jury size. *Law and Human Behavior, 21, 451-467.*

Schachter, S. (1959). *The psychology of affiliation.* Stanford, CA: Stanford University Press.

Schaie, K. W. (1986). Beyond calendar definitions of age, time, and cohort: The general develop-mental model revisited. *Developmental Review, 6,* 252-277.

Schlenger, W. E., Caddell, J. M., Ebert, L., Jordán, B. K., Rourke, K. M., Wilson, D., Thajn, L., Dennis, J. M., Fairbank, J. A. y Kulka, R. A. (2002). Psychological reactions to terrorist attacks: Fin-dings from the National Study of Americans' Reactions to September 11. *Journal of the American Medical Association, 288,* 581-588.

Schoeneman, T J. y Rubanowitz, D. E. (1985). Attributions in the advice columns: Actors and observers, causes and reasons. *Personality and Social Psychology Bulletin, 11,* 315-325.

Schwarz, N. (1999). Self-reports: How the ques-tions shape the answers. *American Psychologist, 54,* 93-105.

Scribner, S. (1997). Studying literacy at work: Bringing the laboratory to the field. En E. Tor-bach, R. J. Falmagne, M. B. Parlee, L. M. W. Martin y A. S. Kapelman (Eds.), *Mind and social practice: Selected writings of Sylvia Scribner.* Cambridge: Cambridge University Press.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515-530.

Sebald, H. (1986). Adolescents' shifting orientation toward parents and peers: A curvilinear trend over recent decades. *Journal of Marriage and the Family, 48,* 5-13.

Shadish, W. R., Cook, T D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Shepard, R. N. y Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

Sidman, M. (1960). *Tactics of scientific research.* Nueva York: Basic Books.

Sieber, J. E. (1992). *Planning ethically responsible research: A guide for students and internal review hoards.* Newbury Park, CA: Sage.

Sieber, J. E., Iannuzzo, R. y Rodriguez, B. (1995). Deception methods in psychology: Have they changed in 23 years? *Ethics and Behavior, 5,* 67-85.

Siegel, S. y Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* Nueva York: McGraw-Hill.

Sigall, H. y Mills, J. (1998). Measures of independent variables and mediators are useful in social psychology experiments. But are they necessary? *Personality and Social Psychology Review, 2,* 218-226.

Silverman, I. (1975). Nonreactive methods and the law. *American Psychologist, 30,* 764-769.

Silverman, I. y Margulis, S. (1973). Experiment title as a source of sampling bias in commonly used "subject-pool" procedures. *Canadian Psychologist, 14,* 197-201.

Singh, D. (1993). Adaptive significance of female . physical attractiveness: Role of waist-to-hip ra-tio. *Journal of Personality and Social Psychology, 65,* 293-307.

Skinner, B. F. (1953). *Science and human behavior.* Nueva York: Macmillan.

Smart, R. (1966). Subject selection bias in psychological research. *Canadian Psychologist, 7,* 115-121.

Smith, C. P. (1983). Ethical issues: Research on deception, informed consent, and debrieñng. En L. Wheeler y P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4). Newbury Park, CA: Sage.

Smith, M. L. y Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752-760.

Smith, R. J., Lingle, J. H. y Brock, T. C. (1978). Reactions to death as a function of perceived similarity to the deceased. *Omega, 9,* 125-138.

Smith, S. M. y Shaffer, D. R. (1991). Celerity and cajolery: Rapid speech may promote or inhibit persuasión through its impact on message elaboration. *Personality and Social Psychology Bulletin, 17,* 663-669.

Smith, S. S. y Richardson, D. (1983). Amelioration of harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology, 44,* 1075-1082.

Smith, S. S. y Richardson, D. (1985). On deceiving ourselves about deception: A reply to Rubin. *Journal of Personality and Social Psychology, 48,* 254-255.

Smith, V. L. y Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology, 72,* 294-300.

Solomon, R. L. (1949). An extensión of control group design. *Psychological Bulletin, 46,* 137-150.

Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology, 51,* 709-725.

Steele, K. M., Bass, K. E. y Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science, 10,* 366-369.

Steinberg, L. y Dornbusch, S. M. (1991). Negative correlates of part-time employment during adolescence: Replication and elaboration. *Developmental Psychology, 27,* 304-313.

Stephan, W. G. (1983). Intergroup relations. En D. Perlman y P. C. Cozby (Eds.), *Social psychology.* Nueva York: Holt, Rinehart & Winston.

Sternberg, R. J. (1993). *The psychologists companion: A guide to scientific writing for students and researchers.* Cambridge: Cambridge University Press.

Stevenson, H. W. y Alien, S. (1964). Adult performance as a function of sex of experimenter and sex *oí subject. Journal of Abnormal and Social Psychology, 68,* 214-216.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N. y Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences, 99(17),* 11531-11536. Consultado en noviembre 1,2002, de http://www.pnas.org/cgi/content/full/99/17/_11531

Sullivan, D. S. y Deiker, T. E. (1973). Subject- experimenter perceptions of ethical issues in human research. *American Psychologist, 28,* 587-591.

Terman, L. M. (1925). *Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children.* Stanford, CA: Stanford University Press.

Terman, L. M. y Oden, M. H. (1947). *Genetic studies of genius: Vol. 4. The gifted child grows up: Twenty-five years' follow-up of a superior group.* Stanford, CA: Stanford University Press.

Terman, L. M. y Oden, M. H. (1959). *Genetic studies of genius: Vol. 5. The gifted group in mid-life: Thirty-five y ears' follow-up of the superior child.* Stanford, CA: Stanford University Press.

Thomas, G. V. y Blackman, D. (1992). The future of animal studies in psychology. *American Psychologist, 47,* 1678.

Thombs, D. L. (2000). A retrospective study of DARÉ: Substantive effects not detected in un-dergraduates. *Journal of Alcohol and Drug Education, 46,* 27-40.

Thompson, W. F., Schellenberg, E. G. y Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science, 12,* 248-251.

Trochim, W. M. (2000). *The Research Methods Knowledge Base* (2a. ed.). Consultado en marzo 6,2003, http://trochim.human.cornell.edu/kb/ index.htm

Tucker, W. H. (1997). Re-considering Burt: Beyond a reasonable doubt. *Journal of the History of the Behavioral Sciences, 33,* 145-162.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning information.* Cheshire, CT: Graphics Press.

U.S. Department of Justice. (1999). *Eyewitness evidence: A guidefor law enforcement.* Consultado en julio 10, 2002, de http://www.ncjrs.org/pdffilesl/nij'/178240.pdf

Vincent, P. C, Peplau, L. A. y Hill, C. T. (1998). A longitudinal application of the theory of reasoned action to women's career behavior. *Journal of Applied Social Psychology, 28,* 761-778.

Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral Science, 11,* 105-114.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, R. y Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2a. ed.). Boston: Houghton Mifflin.

Weber, R. P. (1990). *Basic contení analysis* (2a. ed.). Newbury Park, CA: Sage.

Wells, G. L. (2001). Eyewitness lineups: Data, theory, and policy. *Psychology, Public Policy, and Law, 7,* 791-801.

Wells, G. L., Small, M., Penrod, S. J., Malpass, R. S., Fulero, S. M. y Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647.

Wilkinson, L. y the Task Forcé on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Wilson, D. W. y Donnerstein, E. (1976). Legal and ethical aspects of nonreactive social psy-chological research. *American Psychologist, 31, 765-773.*

Wilson, W H., Ellinwood, E. H., Mathew, R. J. y Johnson, K. (1994). Effects of marijuana on performance of a computerized cognitive-neuromotor test battery. *Psychiatry Research, 51,* 115-125.

Winograd, E. y Soloway, R. M. (1986). On forget-ting the location of things stored in special places. *Journal of Experimental Psychology: General, 115,* 366-372.

Yin, R. K. (1994). *Case study research: Design and methods.* Newbury Park, CA: Sage.

Zimbardo, P. G. (1973). The psychological power and pathology of imprisonment. En E. Aronson y R. Helmreich (Eds.), *Social psychology.* Nueva York: Van Nostrand.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level ofarousal.* Hillsdale, NJ: Erlbaum.

Créditos

Capítulo 2 p. 28, ventana del navegador Netscape © 2002 Netscape Communications Corporation. Reproducido bajo permiso. Netscape Communications no ha autorizado, subvencionado, promocionado ni aprobado esta publicación y no es responsable de su contenido; reproducido bajo permiso de la American Psychological Association, editor de la base de datos PsycINFO, todos los derechos reservados. Prohibida se reproducción sin permiso. **Capítulo 7** p. 129, Fig. 7.1, "Negative Correlates of Part-time Employment During Adolescence," por L. Steinberg y S. M. Dornbusch, 1991, *Developmental Psychology, 27,* pp. 304-313. Copyright © 1991, American Psychological Association. Reproducido bajo permiso. **Capítulo 8** p. 164, Fig. 8.2, reproducido bajo permiso de "Mental Rotation of Three-Dimensional Objects," por R. N, Shepard y J. Metzler, 1971, *Science, 171,* pp. 701-703. Copyright © 1971 American Association for the Ad-vancement of Science. **Capítulo 11** p. 224, Fig. 11.2 y p. 225, Fig. 11.3, "Reforms as Experiments," por D. T. Campbell, 1969, *American Psychologist, 24,* pp. 409-429. Copyright © 1969 por la American Psychological Association. Reproducido bajo permiso. **Apéndice C** p. 396, Tabla C.2, tabla adaptada de Fisher y Yates, *Statistical Tables for Biological, Agricultural, and Medical Research* (1974, 6a. ed.), Londres: Longman. Copyright © 1963 por R. A. Fisher y F. Yates. Reproducido bajo permiso de Pearson Education Limited.

índice analítico

A

Abreviaturas, 327-329
Aleatorización, 81-82
Alfa de Cronbach, 98
Alternativas de respuesta, 134, 136-137
American Psychological Association (APA)
  Código de Ética, 40, 56-60, 61, 63
  Manual de Publicaciones, 18, 296, 301
  programa *APA-Style Helper*, 301 recursos de información, 18, 24, 33, 296, 299
  reuniones profesionales, 193 sitio de Internet, 299, 302
American Psychological Society (APS), 193, 299
Análisis de contenido, 123-124
Análisis de covananza, 167
Análisis de resultados de investigación, 236-237
  comparación de medias grupales, 237
  comparación de porcentajes grupales, 236-237
  correlación de puntajes individuales, 237
  meta-análisis para, 296-298
  programas computacionales para, 278-280
Análisis de riesgo-beneficio, 40, 62-63
Análisis de ruta, 254
Análisis de varianza, 247, 281, 374-385
  para dos variables independientes, 378-383
  para el diseño de medidas repetidas, 383-389
  para una variable independiente, 374-378
  véase también prueba *F*
Análisis del caso negativo, 116
Annual Review of Psychology, 33
APA. *Véase American Psychological Association*
Apéndice de reportes, 314
Aprobación social, 130
APS. *Véase American Psychological Society*
Archivos de encuesta, 122
Aristóteles, 5
Artículos. *Véase* Artículos de investigación
Artículos de investigación
  de revisión de la literatura, 33
  resúmenes de, 24, 27-32
  secciones principales de los, 34-36
  véase también Reportes de investigación
Asignación aleatoria, 160, 390
Aspectos éticos, 39-65
  análisis de riesgo-beneficio y, 40-43, 62-63
  Código de ética de la APA y, 56-60, 61, 63
  compromisos del investigador y, 54
  consentimiento informado y, 43-49
  el *Reporte Belmont* y, 40
  engaño y, 46-49
  experimento de Milgram y, 39-40
  falsificación, 63-65
  investigación con animales y, 60-61
  manipulación de variables y, 181
  método experimental y, 86, 185
  principio de justicia y, 53-54
  proceso de indagación y, 49-50, 192
  reglamentos federales y, 54-56
Asuntos de privacidad, 42-43
Autonomía, 44-46
Autoridad, 5

B

Base de datos *PsycARTICLES*, 33
Base de datos *PsycFIRST*, 24
Base de datos *PsycINFO*, 24
  realización de búsquedas en, 24, 27-32
  resumen de estrategias de búsqueda en, 31
Beneficio, 40
Bloque de ensayos, 164
Bloques aleatorizados, 164-165
Breuning, Stephen, 64
Broum vs Board ofEducation (1954), 3
Burt, Sir Cyril, 64, 65
Búsqueda por medio de comodines, 31

C

Campbell, Donald, 215
Características de la demanda, 51, 186
Características de personalidad, 124-125
Caso del listo Hans, el, 188
  CATI (por sus siglas en inglés), entrevista telefónica por computadora (SCET), 141

Causa necesaria, 85
Causa suficiente, 85
Causalidad
dirección de causa y efecto, 79
requisitos para establecer, 9-10, 84-85
Center for Mental Health Services, 299
Ciencia
metas de la, 8-10
teoría desarrollada de la, 6
Cifras en los reportes, 313, 316, 365-366
Coeficiente de confiabilidad, 97
Coeficientes de correlación, 75, 243-247, 387
Coeficiente de correlación producto-momento de Pearson (r de Pearson), 97, 243-246, 387
ejemplo del uso del comando, 387-385
relaciones curvilineales y, 247
restricción del rango y, 246-247
significancia del, 277, 279
tabla de valores críticos, 401
Coerción, 46 Cómplices, 178
Comportamiento
causas del, 9-10
descripción del, 8, 87-88, 111
explicaciones del, 10
predicciones del, 9, 88
Compromisos del investigador, 54
Computadoras
análisis de datos en las, 278-280
análisis de preguntas de encuestas en, 133
asignación con números aleatorios de, 390
programa de estilo de redacción de la APA para, 301
sistema de entrevista telefónica para, 141
véase también Internet
Confiabilidad, 95-100
consistencia interna, 98-99
entre jueces, 99
por mitades, 98
prueba-reaplicación (test-retest), 97-98
Confiabilidad de consistencia interna, 98-99
Confiabilidad de la prueba-reaplicación, 97-98
Confiabilidad entre jueces, 99
Confiabilidad por mitades, 99
Confidencialidad, 42-43
Conjunto de respuestas, 130
Conjunto de respuestas de "decir sí", 133
Conjunto de respuestas "que dicen no", 133
Consentimiento, 43
Consentimiento informado, 43-49
aspectos de autonomía y, 44, 46
Código de ética y, 57-58
engaño y, 46-48
forma empleada para, 44
temas de información y, 46
Consideraciones culturales, 289-290
Consideraciones del género, 286
Consideraciones gramaticales, 329-331
palabras/frases de transición, 329-330
palabras/frases problemáticas, 330-331
Contrabalanceo, 163-165
Control experimental, 81
Controles, 185-190
diseños de grupo control no equivalente, 222-224
expectativas de los participantes y, 186-187
expectativas del experimentador y, 187-190
Correlación, 386
espuria, 4
múltiple, 250-252
parcial, 252 Correlaciones del reactivo con la puntuación total, 99
Costos
de la manipulación de variables, 181
de medidas, 185
Covariación de causa y efecto, 9, 84
Cuadrados latinos, 163-164, 402-403
Cuadrados medios, 376-377, 382
Cuestionarios
aplicación de, 138-140
formato de, 137
reactivos de relleno en, 186
refinamiento de las preguntas de, 138

D

d de Cohén, 268
Daño físico, 41
Darwin, Charles, 20
Datos
análisis por computadora de, 278-280
fabricación de, 63-64
Decisiones correctas, 271
Definiciones operacionales, 71-72
Descripción del comportamiento, 8, 87-88
Desviación estándar, 244
Diagramas de dispersión, 244-247
Diferencias de selección, 156, 222-223
Difusión de la responsabilidad, 70-71
Dirección de causa y efecto, 79
Diseño ABA, 226-227
Diseño de cuatro grupos de Solomon, 158, 159, 291
Diseño de experimentos. *Véase* Diseño experimental
Diseño de grupo control no equivalente, 222-223
Diseño de grupos apareados, 166-167
Diseño de grupos independientes, 159, 160
diseño de medidas repetidas vs., 165-166
diseños factoriales y, 208 Diseño de línea base múltiple, 228 Diseño de medidas repetidas, 159, 160-166

análisis de varianza para, 383-385
diseño de grupos independientes contra, 165-166
diseños factoriales y, 208
intervalos de tiempo y, 165
técnicas de contrabalanceo para, 163-165
ventajas y desventajas del, 161-162
Diseño de retiro del tratamiento, 226
Diseño de serie de control, 225-226
Diseño de series de tiempo interrumpidas, 224-225
Diseño de un grupo con pretest-postest, 219-222
Diseño de un grupo con únicamente postest, 218-219
Diseño de únicamente postest, 155-156, 158
de un grupo, 218
Diseño ex post facto, 87
Diseño experimental, 153-171
diseño de grupos apareados, 166-167
diseño de grupos independientes, 159, 160
diseño de medidas repetidas, 159, 160-166
diseño de pretest-postest, 156-157, 158
diseño de únicamente postest, 155-156, 158
diseños complejos, 197-212
diseños cuasiexperimentales, 215, 218-226
diseños de investigación del desarrollo, 167-171
diseños experimentales de caso único, 226-230
diseños factoriales, 200-212
Diseño factorial 2x2, 200
efectos principales del, 201-202
resultados del, 205-207
Diseño factorial 2 x 2 x 2 , 210-212
Diseño factorial 2x3, 209-210
Diseño factorial mixto, 208, 209
Diseño placebo balanceado, 186-187
Diseño pretest-postest, 156-157, 158
de grupo control no equivalente, 223-224
de un grupo, 219-222
Diseño reversible, 226-228
Diseños cuasiexperimentales, 215, 218-226
diseño de grupo control no equivalente, 222-223
diseño de series de control, 225-226
diseño de series de tiempo interrumpidas, 224-225
diseño de un grupo con pretest-postest, 219-222
diseño de un grupo con únicamente postest, 218-219
diseño pretest-postest de grupo control no equivalente, 223-224
Diseños de investigación del desarrollo, 167-171
comparación de, 169-171
método longitudinal, 169
método secuencial, 168, 171
método transversal, 168 Diseños de un solo sujeto, 226
Diseños experimentales de un solo sujeto, 226-230
diseño reversible, 226-228
diseños de línea base múltiple, 228-229
réplicas en los, 229-230
Diseños factoriales, 200-212
diseño mixto, 208, 209
diseño IV x PV, 203-204
efectos principales simples y, 207-208
explicación de, 200
incremento de la complejidad de, 209-212
interpretación de, 200-203
procedimientos de asignación y, 208-209
resultados de un diseño 2x2, 205-207
variables moderadoras, 204-205
Distribución binomial, 262
Distribuciones de frecuencias, 237-240
Distribuciones muéstrales, 262-263

E

Ecuaciones de regresión, 249-250, 251
Efecto de contraste, 162
Efecto de fatiga, 162
Efecto de influencia, 162
Efecto de la práctica, 162
Efecto de piso, 183
Efecto de techo, 183
Efecto Mozart, 294
Efectos de cohorte, 170
Efectos de expectativa, 187-190
investigación sobre los, 188-189
soluciones para los, 189-190
Efectos de la historia, 219
Efectos de la maduración, 220
Efectos de la prueba, 220
Efectos del orden, 162 Efectos
placebo, 186-187
Efectos principales, 200, 201
simples, 207-208
Efectos principales simples, 207-208
Electroencefalograma (EEG), 182, 183
Electromiograma (EMG), 182, 183
Empirismo, 5
Encabezados de los reportes, 317-318
Encuadre de la muestra, 148
Encuestados, 18
Encuestas, 127-151
aplicación de, 138-141
cuestionarios para, 137-140
definición de objetivos de investigación para, 130-131
elaboración de preguntas de, 130-133
entrevistas para, 140-141
escalas de puntuación para, 134-135
estudio de cambios en el tiempo con, 142
proceso de muestreo para, 142-151
razones para realizar, 128-130
respuestas a preguntas de, 133-137
Encuestas por correo, 139

Engaño, 46-49, 59
Entrevistador, sesgo del, 140
Entrevistas, 140-141
Entrevistas cara a cara, 141
Entrevistas telefónicas, 141
Error
  aleatorio, 161
  de medición, 95, 96
  de muestreo, 143
  tipo I y tipo II, 270-273
Errores tipo I, 271, 272, 274
Errores tipo II, 271-275
Escala de alta frecuencia, 137
Escala de búsqueda de sensaciones, 103-104
Escala de intervalo, 106-107, 235-236, 280
Escala de puntuación gráfica, 135
Escala no verbal para niños, 135
Escalas de diferencial semántico, 135
Escalas de medición, 105-107, 235-236
  de intervalo, 106-107
  de razón,106-107
  importancia de, 107
  nominal, 105-106
  ordinal, 106
Escalas de puntuación, 134-135
Escalas de razón, 107, 235, 236, 280
Escalas nominales, 105-106, 235, 280
Escalas ordinales, 106, 235, 236, 280
Escepticismo, 5 Estadística, 235
  descriptiva, 240-241
  inferencial, 258-259
  reporte estadístico, 331
Estadística diferencial, 258-259
Estadísticas descriptivas, 240-241, 367-370
  medidas de tendencia central, 240-241
  medidas de variabilidad, 241, 369-370
Estilo de citas, 318-321
Estilo de redacción, 302-306
  claridad del, 302-303
  gramática y, 329-331
  lenguaje prejuiciado y, 304-305
  requisitos de estilo de la APA, 306
  voz pasiva contra activa, 303-304
Estrategias de búsqueda, 31-32
Estrés, 41-42
Estudiantes universitarios, 285, 288-289
Estudio de Sífilis de Tuskegee, 53-54
Estudios de caso, 119-121
Estudios de panel, 142
Estudios de simulación, 51-52
Estudios piloto, 123
Etnografía, 123
Evaluación
  de muestras, 148-151
  de programas, 12, 215-218
Evaluación de la eficiencia, 217
Evaluación de la teoría del programa, 215, 216
Evaluación de necesidades, 216
Evaluación del proceso, 217
Evaluación del resultado, 217
Expectativas
  del experimentador, 187-190
  del participante, 186-187
Experimentador, sesgo del, 187
Experimento de Asch sobre el conformismo, 179
Experimento de la prisión de Stanford, 51
Experimento de Milgram, 39-40, 46-47
Experimentos
  análisis de resultados de, 193
  artificialidad de, 85-86
  asignaciones de participantes a los, 159
  aspectos de control en, 185-190
  características de demanda, 186
  efectos de las expectativas en, 187-190
  efectos placebo en, 186-187
  factor de abandono en, 157
  preparación del terreno, 176
  selección de participantes para, 175-176
  validez interna de, 155
  variable dependiente en, 181-185
  variable independiente en, 155, 176-181
  variables intervinientes en, 154-155
Experimentos ciegos simples, 189
Experimentos de campo, 86, 292, 293
Experimentos doble ciego, 190
Experimentos honestos, 52-53
Explicación de comportamiento, 10
Explicaciones alternativas, 10, 84

F

Falsificación, 63-65
Falla de un instrumento, 220
Fraude, 63-65
Fuentes electrónicas, 326-327
Fuentes secundarias, 325

G

Galton, Sir Francis, 105
Generalización, 285-299
  características del experimentador y, 290-291
  como interacción estadística, 287-288
  consideraciones culturales y, 289-290
  evaluación de, 296-298

evidencia para, 298-299
investigación de laboratorio y, 292-293
poblaciones de investigación y, 285-289
pretest y, 291
réplica y, 293-296 Generalización del experimentador, 290-291
Gmliano, Traci A., 333
Grados de libertad, 266, 373
Gradúate Record Exam (GRE), 102
Gráficas
de distribuciones de frecuencias, 238-240
de relaciones entre variables, 74, 241-242
Gráficas de barras, 238, 239
Gráficas de pastel, 238, 239
Grupos
aplicación de encuestas a, 138-139
placebo, 186-187
Grupos de enfoque, 141-142

H

Health and Human Services (HHS), 54
Hipótesis, 17
Hipótesis de investigación, 259
Hipótesis nula, 259-260
errores tipo I y tipo II y, 271-273
resultados no significativos y, 274-276
Human Relations Área Files (HRAF), 123

I

Ideas, fuentes de, 18-23
Imagen de resonancia magnética (IRM), 183
Incertidumbre, reducción de la, *76*
Indagación, 49-50, 59-60, 192
Información de hechos, 131
Informantes, 18
Institutional Animal Care and Use Committee (CICUA), 61
Institutional Review Board (CIR), 54-56
Interacciones, 201, 202-203
efectos principales simples y, 207-208
estadísticas, 287
variables moderadoras y, 204-205
Internet
búsqueda de información en, 33-34
cita de fuentes de, 326
encuestas realizadas por medio de, 139
recursos psicológicos en, 299
Inter-university Consortium for Political and Social Research (ICPSR), 122 Intervalos de confianza, 143-144, 268-269 Intuición, 4
Inventario de personalidad NEO (IP-NEO), 124
Investigación
análisis de resultados de, 193, 236-237
básica y aplicada, 11-14
bibliográfica, 23-36
de laboratorio, 292-293
documental, 121-124
mejoría de la vida por medio de la, 298-299
método científico de, 3-8
normas éticas para la, 56-60
preparación del terreno para, 176
reportes de, 193, 301-366
selección de participantes para, 175-176
usos y aplicaciones de la, 2-3, 298-299
Investigación anterior, 22-23
Investigación aplicada, 11-13
Investigación básica, 11, 13-14
Investigación bibliográfica, 23-36
artículos de investigación y, 34-36
búsqueda en Internet y, 33-34
psychological abstraéis y, 24
recursos en bases de datos y, 24, 27-32
revisiones de la literatura y, 33
revistas científicas y, 23, 25-27
Investigación con animales, 60-62
Investigación de laboratorio, 292-293
Investigación de riesgo mínimo, 55
Investigación documental, 121-124
análisis de contenido y, 123-124
archivos de encuestas y, 122
registros escritos y, 122-123
registros estadísticos y, 121-122
Investigación exenta, 55
Investigación psicológica
mejoría de la vida por medio de la, 298-299
normas éticas para la, 56-60
redacción de reportes sobre la, 193, 301-366
riesgos potenciales en la, 41-43
véase también Investigación IRM funcional (fíRM, por sus siglas en inglés), 183
IV x PV, diseños, 203-204

J

Journal of Abnormal and Social Psychology, 285
Journal of Experimental Psychology, 285
Journal of Personality and Social Psychology, 48-49
Justicia, 53-54

K

Kappa de Cohén, 99
Knight, Jennifer L,, 333

L

Law School Admission Test (LSAT), 102
Lenguajes prejuiciados, 304-305
Levine, Robert, 20
Línea base, 226
Lincamientos para la conducta ética en el cuidado y uso animales (APA), 62
Lineamientos para los reportes, 302
Lista de referencias, 314, 321-327, 356-362
Localidad geográfica (lugar), 287

M

Magnitud del efecto, 247-248
  cálculo del, 268
  estadístico chi cuadrada y, 386
  estadístico *F* y, 386-387
Manipulación de eventos, 178
Manipulación de variables, 176-181
  costo de la manipulación, 181
  fuerza de la manipulación, 180-181
  tipos de manipulación, 176-179
Manipulación por etapas, 178-179
Manipulaciones directas, 176-178
Manual de Publicaciones de la Asociación Psicológica Americana, 18, 296, 301
Margen del error, 143
Media, 240, 368
Mediana, 240, 368
Medición, 95
Medidas
  autorreporte de, 182
  conductuales, 182
  confiabilidad de las, 95-100
  costo de las, 185
  de tendencia central, 367-369
  de variabilidad, 369-370
  fisiológicas, 182-183
  múltiples, 184-185
  precisión de las, 99-100
  reactividad de las, 104-105
  sensibilidad de las, 183-184
  validez de constructo de las, 100-104
Medidas de autorreporte, 182
Medidas no reactivas, 104
Medidas sin obstrucción, 104
Mental Measurements Yearbook (Conoley y Kramer) 125
Meta-análisis, 296-298
Metas de la ciencia, 8-10
Método científico, 3-8
Método correlacional, 78
Método cualitativo, 111, 113
Método cuantitativo, 111, 114
Método experimental, 78, 154
  características del, 80-82
  desventajas del, 85-89
Método longitudinal, 168-171
Método no experimental, 78-80
Método secuencial, 168, 171
Método transversal, 168, 169-171
Métodos múltiples, 88-89
Miller, George, 298 Moda, 240, 368
Modelos estructurales, 253-254
Mortalidad, 157
Muestra aleatoria, 145, 390
Muestra homogénea, 246
Muestreo, 119, 145-151
  evaluación de muestras, 148-151
  intervalos de confianza y, 143-144
  no probabilístico, 145, 146-147
  probabilidad de, 145-147
  tamaño de la muestra y, 144, 263, *276*
  técnicas de, 145-147
Muestreo accidental, 147
Muestreo aleatorio estratificado, 145-146
Muestreo de conveniencia, 149-151
Muestreo no probabilístico, 145, 146-147
Muestreo por cuotas, 147
Muestreo por racimos, 146
Muestreo probabilístico, 145-146

N

Niños
  escala de puntuación no verbal para, 135
  investigación ética con, 44, 46
Nivel alfa, 262
Notas al pie de página en los reportes, 315, 364
Notas del autor en reportes, 315, 363

O

Observación, 19-20, 43, 111-119
Observación de campo, 112
Observación naturalista, 112-115
  aspectos en la, 114-116
  límites de la, 115-116
Observación oculta, 114
Observación sistemática, 116-119
  aspectos metodológicos en la, 118-119
  sistemas de codificación en la, 117-118
Operadores boléanos, 30

p

Palabras/frases de transición, 329-330
Palabras/frases problemáticas, 330-331
Participantes, 18
  asignación aleatoria de los, 390
  control de las expectativas de los, 186-187
  generalización de los resultados de los, 285-289
  observación naturalista de los, 114-115
  proceso de selección de los, 175
Pavlov, I van, 20
Pensamiento práctico, 112
Piaget, Jean, 87
Placebo, grupos, 186-187
Plagio, 63-65
Población, 142
Polígonos de frecuencias, 239-240
Portada de los reportes, 307, 334
Potencia, 276
Precedencia temporal, 9, 84
Predicción, 9, 17, 88
Preguntas
  elaboración para encuestas, 130-133
  importancia de la redacción de las, 131-133
  respuestas a las, 134-137
Preguntas abiertas, 133-134
Preguntas de respuesta cerrada, 133-134
Preguntas dobles, 132
Preguntas sesgadas, 132
Pretest, 291
Principios éticos de los psicólogos y Código de conducta (APA), 56
Probabilidades, 260-262
Problema de la tercera variable, 79-80, 252
Problemas prácticos, 23
Procedimiento de "pensar en voz alta", 138, 191
Programa DARÉ, 215, 225
Programa de cómputo QUAID, 133
Programas de hoja de cálculo, 390
Programas estadísticos de cómputo, 278
Propuestas de investigación, 190-191
Prueba chi cuadrada, 280, 281
  magnitud del efecto y, 386
  significancia estadística y, 371-373
  tabla de valores críticos, 396
Prueba de habilidades administrativas, 100, 102
Prueba de una cola, 267
Prueba del signo, 280
Prueba F, 264, 267-268
  magnitud del efecto y, 386
  significancia estadística y, 374, 386
  tabla de valores críticos, 398-400
  véase también Análisis de varianza
Prueba *H* de Kruskal-Wallace, 281
Prueba *Shy* Q, 100, 103
Prueba *t*, 263-268, 280, 281
  grados de libertad para la, 266
  tabla de valores críticos, 397
Prueba Tde Friedman, 281
Prueba Tde Wilcoxon, 280
Prueba *U* de Mann-Whitney, 280
Pruebas de dos colas, 267
Pruebas estadísticas, 367-389
  correlación y magnitud del efecto, 386-389
  estadística descriptiva, 367-370
  significancia estadística, 370-386
Psicobiografía, 119
Psicología
  principales revistas científicas de, 25-27
  resumen de los artículos en, 24, 27-32
PsycARTICLES, base de datos, 33
PsycFIRST, base de datos, 24
Psychological Abstracts, 24
Psychological Bulletin, 33
PsycINFO, base de datos, 24
  realización de búsquedas en, 24, 27-32
  resumen de estrategias de búsqueda en, 32
Puntuación verdadera, 95

R

Rango, 241, 369
Reactividad, 104, 118
Reactivos de relleno, 186
Realismo, 292-293
Realismo experimental, 292-293
Realismo mundano, 292-293
Recursos en línea. *Véase* Internet
Referencias enmascaradas, 315
Registros de comunicación masiva, 122-123
Registros escritos, 122-123
Registros estadísticos, 121-122
Registros públicos, 121-122
Regresión estadística, 220-222
Regresión hacia la media, 220-222
Relación curvilínea, 73, 74, 75, 199, 247
Relación lineal positiva, 73
Relaciones curvilineales de U invertida, 199
Relaciones lineales, 73-74
Relaciones lineales negativas, 73, 74
Réplicas, 293
  conceptuales, 295-296
  directas, 293-295
  en los diseños de caso único, 229-230
  importancia de las, 277, 293-296
Reporte Belmont, el, 40

Reportes de investigación, 301-366
abreviaturas utilizadas en los, 327-329
cifras y estadísticas en los, 331-332
consideraciones gramaticales en los, 329-331
encabezados utilizados en los, 317-318
estilo de citas para los, 318-321
estilo de lista de referencias para los, 321-327
estilo de redacción de los, 302-306
lincamientos del formato para los, 305-306
muestra de un reporte completo, 333-366
organización de los, 307-317
reconocimiento del trabajo de otros en los, 303
recursos para los, 301
Reportes. *Véase* Reportes de investigación
Respuesta galvánica de la piel (RGP), 182-183
Restricción del rango, 246-247
Resultados no significativos, 274-276
Resúmenes, 24-34, 308-309, 335
Retórica (Aristóteles), 5
Reuniones profesionales, 193
Revisión de colegas, 6, 194
Revisiones de la literatura, 33, 296, 309
Revistas científicas, 23-24
formato para lista de referencias en, 322
lista de las principales, 25-27
publicación de investigaciones en, 193-194
resúmenes de, 24
Riesgos de la investigación psicológica, 41-43
Roles, juego de, 51-52
Rosenthal, Robert, 297
Ruggiero, Karen, 64

S

Science Citation Index (SCI), 32
Scribner, Sylvia, 112
Sección de discusión de los reportes, 36, 313-314, 349-356
Sección de resultados de los reportes, 35, 311-313, 347-349
Sección del método de los reportes, 35, 310-311, 343-3
Sección introductoria de los reportes, 35, 309-310, 336-343
Sensibilidad, 183-184
Sentido común, 18-19
Serendipity, 20
Sesgo
del entrevistador, 140
del experimentador, 187
Significancia estadística, 143, 249, 260, 370-385
análisis de varianza y, 374-385
coeficiente *r* de correlación de Pearson y, 277, 389
interpretación de resultados no significativos, 274-276
niveles de significancia y, 273-274
panorama general de la, 269-270
prueba chi cuadrada y, 371-373
pruebas de, 280-281, 370-385
Significancia. *Véase* Significancia estadística
Simulación entre naciones, 51-52
Sistema de Codificación de Interacción Familiar (SCIF), 118
Sistema SYMLOG de codificación, 118
Sistemas de codificación, 117-118
Skinner, B. R, 14, 226
Sports Illustrated, 221
SSCI *(Social Sciences Citation Index)*, 32
Sujetos. *Véase* Participantes Suma de cuadrados, 374-376

T

Tabla de números aleatorios, 390-395
Tablas en los reportes, 312-313, 315
Tablas estadísticas, 390-401
tabla de números aleatorios, 390-395
valores críticos de chi cuadrada, 396
valores críticos de *F*, 398-400
valores críticos de r, 401
valores críticos de *t*, 397
Tasa de respuesta, 148-149
Tendencia central, 240-241, 367-369
Teoría de la evolución, 21
Teoría desarrollada de la ciencia, 6
Teorías, 20-21
Test Critiques (Keyser y Sweetland), 125
Títulos, 307-308
Trastorno por estrés postraumático (TPEPT), 53

V

Validez, 89
aparente, 100-101
concurrente, 103
convergente, 103
de constructo, 89-90, 100-104
discriminante, 103-104
interna, 90, 155
orientada al criterio, 101
predictiva, 102
Validez de constructo, 89-90, 100-104
indicadores de, 100-102
Validez de convergencia, 103

Validez externa, 90, 285
Validez interna, 90, 155, 285
Validez predictiva, 102
Valores críticos
  de chi cuadrada, 396
  de F, 398-400
  de r, 401
  de *t*, 397
Variabilidad, 241, 369-370
Variabilidad aleatoria, *76-77*
Variable de criterio, 249
Variable predictora, 250
Variables, 70-77
  categorías de, 70
  de criterio, 249
  definiciones operacionales de, 71-72
  dependientes, 82-83
  escalas de medición y, 105-108
  independientes, 82-83, 176-181
  intervinientes, 80, 154-155
  moderadoras, 204
  predictoras, 250
  relaciones entre, 72-77, 241-242
Variables de respuesta, 70
Variables del participante, 70, 87
Variables del sujeto, 70
Variables dependientes, 83
  mediciones en experimentos, 181-185
  sensibilidad de, 183-184
  véase también Variables independientes
Variables independientes, 83
  análisis de varianza para, 374-383
  diseño experimental y, 155
  diseños factoriales y, 200-212
  efectos principales de, 200, 201-202
  incremento de la complejidad de, 198-199, 209-212
  interacciones entre, 201-203
  manipulación de, 176-181
  véase también Variables dependientes
Variables intervinientes, 80, 154-155
Variables moderadoras, 204
Variables situacionales, 70
Varianza, 241, 267-268, 369-370
Varianza del error, *76-77*, 268
Varianza dentro de grupos, 268
Varianza entre grupos, 268
Varianza sistemática, 267
Verificación de la manipulación, 191-192
Voluntarios, 286
Voz activa, 303-304
Voz pasiva, 303-304

W

World Wide Web. *Véase* Internet

Cozby

Métodos de Investigación del Comportamiento

8a. edición

La psicología científica comparte características tanto de ciencias como la biología y la química, como de las ciencias sociales. Por lo tanto, los métodos de investigación que se emplean en psicología son comunes con los que se aplican en otras áreas del comportamiento. Como profesión, la psicología abarca un vasto campo de aplicaciones, desde las más tradicionales, como la psicología clínica y la educativa, hasta las más recientes, como la psicología ambiental y la psicología de la política. Esta obra presenta todo el panorama de métodos de investigación comúnmente empleados en la psicología de manera sencilla y práctica, con ejemplos de la vida cotidiana.

Algunos de los temas que se exponen en este texto son:

- Comprensión científica del comportamiento.
- Cómo iniciar una investigación.
- Ética de la investigación.
- Compromisos del investigador.
- Estudio del comportamiento.
- Conceptos sobre medición.
- Observación del comportamiento.
- Investigación de encuestas.
- Diseño experimental.
- Realización de experimentos.
- Comprensión de los resultados de la investigación.

El libro incluye ejercicios fáciles de realizar de manera autodidacta, además de cubrir de manera adecuada la gama de técnicas y procedimientos de método que se utilizan en la investigación básica y aplicada en psicología. Es un texto actualizado que incorpora recursos didácticos de vanguardia, como la indicación de los conceptos relevantes, las preguntas de repaso, las actividades prácticas y un sitio en Internet en el que los alumnos pueden comprobar sus avances, resolver sus dudas y desarrollar sus ejercicios.

www.mhhe.com/sociales/cozby8e

McGraw-Hill Interamericana

The McGraw-Hill Companies

ISBN 970-10-4825-3

9 789701 048252

Visita nuestra página WEB
www.mcgraw-hill.com.mx